This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

OM of: US-09-673-254-1 to: Issued_Patents_AA:* out_format : pfs

Date: Jun 11, 2002 8:08 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09673254/runat_11062002_114211_919/app_query.fasta_1.2973
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
[redacted]APEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
[redacted]RIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
[redacted]_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=7 -MODE=LOCAL
[redacted]FMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09673254_@CGN1_1_51 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-673-254-1
Query length: 2870
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 38.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-396-218-2	-	2209.00	2304.70	1.6e-121	422	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-760-116-2	-	2209.00	2304.70	1.6e-121	422	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-982-9	-	2104.50	2195.02	1.9e-115	443	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-653-650A-9	-	2104.50	2195.02	1.9e-115	443	!	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-982-8	-	2093.50	2182.90	8.5e-115	474	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-653-650A-8	-	2093.50	2182.90	8.5e-115	474	!	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-982-5	-	2076.00	2165.68	8.7e-114	422	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-653-650A-5	-	2076.00	2165.68	8.7e-114	422	!	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-901-306-2	-	1014.00	1059.17	5.5e-52	287	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-180-271-2	-	1014.00	1059.17	5.5e-52	287	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-901-306-5	-	632.00	668.15	8.2e-30	117	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-5	-	514.00	525.41	8.9e-23	960	!	S
[redacted]2_6/ptodata/1/iaa/5A_COMB.pep:US-08-642-255-62	+	491.00	500.42	2.0e-21	1064	!	
[redacted]2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-78	-	489.00	500.06	2.5e-21	882	!	
[redacted]2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-79	-	487.00	495.03	3.5e-21	1213	!	
[redacted]2_6/ptodata/1/iaa/5A_COMB.pep:US-08-642-255-62	-	484.00	493.10	5.0e-21	1064	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-78	+	473.50	483.86	2.0e-20	882	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-6	-	471.50	480.99	2.6e-20	960	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-79	+	466.50	473.60	5.4e-20	1213	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-931-820-1	-	462.50	470.69	9.0e-20	1057	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-5	+	439.00	447.01	2.1e-18	960	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-642-255-72	-	437.50	444.49	2.6e-18	1065	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-945-283-2	+	434.50	435.74	4.3e-18	1958	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-18	-	434.00	438.71	4.3e-18	1341	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-570-573-18	-	434.00	438.71	4.3e-18	1341	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-548-608-18	-	434.00	438.71	4.3e-18	1341	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-931-820-4	-	432.50	439.33	5.0e-18	1057	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-642-255-72	+	430.50	437.17	6.6e-18	1065	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-642-255-53	-	428.00	437.47	8.7e-18	777	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-945-283-2	-	427.50	428.42	1.1e-17	1958	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-219-849-4	-	425.50	435.56	1.2e-17	720	!	S
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-21	-	424.00	430.27	1.6e-17	1078	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-570-573-21	-	424.00	430.27	1.6e-17	1078	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-548-608-21	-	424.00	430.27	1.6e-17	1078	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-931-820-1	+	420.50	426.79	2.5e-17	1057	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-21	+	413.50	419.29	6.4e-17	1078	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-570-573-21	+	413.50	419.29	6.4e-17	1078	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-548-608-21	+	413.50	419.29	6.4e-17	1078	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-931-820-4	+	411.50	417.38	8.4e-17	1057	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-19	-	409.50	412.93	1.1e-16	1366	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-570-573-19	-	409.50	412.93	1.1e-16	1366	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-548-608-19	-	409.50	412.93	1.1e-16	1366	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-963-825-20	-	409.50	412.58	1.2e-16	1418	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-010-999-1	-	409.50	412.58	1.2e-16	1418	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-570-573-20	-	409.50	412.58	1.2e-16	1418	!	

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-396-218-2

seq_documentation_block:
; Sequence 2, Application US/08396218
; Patent No. 5695966
; GENERAL INFORMATION:
; APPLICANT: INVENTI, Augusto
; APPLICANT: BREME, Umberto
; APPLICANT: COLOMBO, Anna L
; APPLICANT: HUTCHINSON, Charles R
; APPLICANT: OTTEN, Sharee
; APPLICANT: SCOTTI, Claudio
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIDO, MARMELSTEIN, MURRAY & ORAM
; STREET: 655 Fifteenth Street, N. W., Suite 330 - G
; STREET: Street Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,218
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, Monica C
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P1615-5002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-396-218-2

alignment_scores:
Quality: 2209.00 Length: 422
Ratio: 5.235 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-673-254-1/rev x US-08-396-218-2 ..

Align seg 1/1 to: US-08-396-218-2 from: 1 to: 422

```
    1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ValSerGlyGluAlaProArgValAlaValAspProPheAlaCysProMe 17

    1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

    1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50

1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValLy 67

1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCG 1102
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 sAspProAspLeuAlaProAlaAlaTrpArgGlyValAspAspGlyLeuA 84

1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 GluAlaHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117

1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134

 951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 euThrGluLeuAlaAspAlaSerGlyArgSerGlyLysProAlaGluLeu 150

 901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 IleGlyGlyPheAlaTyrHisPheProLeuLeuValIleCysGluLeuLe 167

 851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValSerValL 184

 801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 euLysAlaLeuGlyLeuGlyGlyProGlnSerGlyGlyGlyAspGlyThr 200

 751 GACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 AspProAlaGlyGlyValProAspThrSerAlaLeuGluSerLeuLeuLe 217

 701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 uGluAlaValHisSerAlaArgArgAsnAspThrProThrMetThrArgV 234

 651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250

 601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrGl 267

 551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284

 501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 laAspAlaAspGluAspAlaValSerArgPheValGluGluAlaLeuArg 300

 451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 TyrHisProProValProTyrThrLeuTrpArgPheAlaAlaThrGluVa 317

 401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 lThrIleGlyGlyValArgLeuProArgGlyAlaProValLeuValAspI 334

 351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350

 301 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 HisProAspArgProSerTrpArgArgLeuThrPheGlyAspGlyProHi 367

 251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384

 201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 lyValLeuArgSerArgPheProGluAlaArgLeuAlaValProTyrAsp 400

 151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GluLeuArgTrpCysArgLysGlyAlaGlnThrAlaArgLeuThrGluLe 417

 101 GCCCGTCTGGCTGCGC 86
     ||||||||||||||||
 417 uProValTrpLeuArg 422
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-760-116-2

```
seq_documentation_block:
; Sequence 2, Application US/08760116
; Patent No. 5786190
;  GENERAL INFORMATION:
;    APPLICANT:  INVENTI, Augusto
;    APPLICANT:  BREME, Umberto
;    APPLICANT:  COLOMBO, Anna L
;    APPLICANT:  HUTCHINSON, Charles R
;    APPLICANT:  OTTEN, Sharee
;    APPLICANT:  SCOTTI, Claudio
;    TITLE OF INVENTION:  PROCESS FOR PREPARING DOXORUBICIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIKAIDO, MARMELSTEIN, MURRAY & ORAM LLP
;      STREET:  655 Fifteenth Street, N. W., Suite 330 - G
;      STREET:  Street Lobby
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/760,116
;      FILING DATE:  3-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,218
;      FILING DATE:  27-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KITTS, Monica C
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-6007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202/638-5000
;      TELEFAX:  202/638-4810
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-760-116-2

alignment_scores:
            Quality: 2209.00              Length:    422
```

```
            Ratio:  5.235                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-673-254-1/rev x US-08-760-116-2   ..

 Align seg 1/1  to: US-08-760-116-2  from: 1  to: 422

    1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ValSerGlyGluAlaProArgValAlaValAspProPheAlaCysProMe 17

    1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

      51 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50

    1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValLy 67

    1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCG 1102
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      67 sAspProAspLeuAlaProAlaAlaTrpArgGlyValAspAspGlyLeuA 84

    1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

    1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GluAlaHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117

    1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134

     951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     134 euThrGluLeuAlaAspAlaSerGlyArgSerGlyLysProAlaGluLeu 150

     901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 IleGlyGlyPheAlaTyrHisProProLeuLeuValIleCysGluLeuLe 167

     851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValSerValL 184

     801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euLysAlaLeuGlyLeuGlyGlyProGlnSerGlyGlyGlyAspGlyThr 200

     751 GACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AspProAlaGlyGlyValProAspThrSerAlaLeuGluSerLeuLeuLe 217

     701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     217 uGluAlaValHisSerAlaArgArgAsnAspThrProThrMetThrArgV 234

     651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250

     601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrGl 267

     551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284

     501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     284 laAspAlaAspGluAspAlaValSerArgPheValGluGluAlaLeuArg 300

     451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TyrHisProProValProTyrThrLeuTrpArgPheAlaAlaThrGluVa 317

     401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     317 lThrIleGlyGlyValArgLeuProArgGlyAlaProValLeuValAspI 334

     351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     334 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350

     301 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 HisProAspArgProSerTrpArgArgLeuThrPheGlyAspGlyProHi 367

     251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384

     201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     384 lyValLeuArgSerArgProProGluAlaArgLeuAlaValProTyrAsp 400

     151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     401 GluLeuArgTrpCysArgLysGlyAlaGlnThrAlaArgLeuThrGluLe 417

     101 GCCCGTCTGGCTGCGC 86
         ||||||||||||||||
     417 uProValTrpLeuArg 422

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-982-9

seq_documentation_block:
; Sequence 9, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/096,982
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
```

```
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  443 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-096-982-9


alignment_scores:
            Quality: 2104.50              Length:    432
              Ratio:  4.987                Gaps:      1
 Percent Similarity:  97.685   Percent Identity: 93.519

alignment_block:
 US-09-673-254-1/rev x US-09-096-982-9    ..

 Align seg 1/1  to: US-09-096-982-9  from: 1  to: 443

    1378 GTACCCCGCGGCGGCGGGCGG...TGCCTCGTGAGCGGCGAGGCGCCCCG 1332
         ||||||   |||:::::::::   ||||||:::|||||||||||||||||
      12 ValProTyrGlyAsnSerGluGlyCysLeuMetSerGlyGluAlaProAr 28

    1331 GGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGC 1282
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
      28 gValAlaValAspProPheSerCysProMetMetThrMetGlnArgLysP 45

    1281 CCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAAC 1232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      45 roGluValHisAspAlaPheArgGluAlaGlyProValValGluValAsn 61

    1231 GCCCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCG 1182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      62 AlaProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaAr 78

    1181 CGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCG 1132
         ||||||||||||||||||||||||||||||||||||||||||||||||:
      78 gGluValLeuAlaAspProArgPheValLysAspProAspLeuAlaProT 95

    1131 CCGCCTGGCGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTG 1082
         ::|||||||||||||||||||||||||||||||||||||||||||||||||
      95 hrAlaTrpArgGlyValAspAspGlyLeuAspIleProValProGluLeu 111

    1081 CGTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCG 1032
         ||||||||||||||||||||||||||||||||   ||||||||||||||||
      112 ArgProPheThrLeuIleAlaValAspGlyGluAspHisArgArgLeuAr 128

    1031 CCGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGG 982
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      128 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrA 145

    981 ATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCC 932
        !||||||||||||||||||   ||||||||||||||||||||||||:::
      145 spArgIleAlaAlaIleAlaAspArgLeuLeuThrGluLeuAlaAspSer 161

    931 TCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCA 882
        |||   |||||||||:::||||||||||||||||||||||||||||||||
      162 SerAspArgSerGlyGluProAlaGluLeuIleGlyGlyPheAlaTyrHi 178

    881 CTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATC 832
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      178 sPheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspP 195

    831 CGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGC 782
        ||||||||||||||||||||||:::|||||||||||||||||||||||||
      195 roAlaMetAlaArgGluAlaValGlyValLeuLysAlaLeuGlyLeuGly 211

    781 GGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCC 732
        ||||||||||||:::|||||||||||||||||||||||||   |||||
      212 GlyProGlnSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 228

    731 GGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCC 682

        ||||||||||||||||||||||||||||||||||||||||||||:::||||
      228 oAspThrSerAlaLeuGluSerLeuLeuLeuGluAlaValHisAlaAlaA 245

    681 GGCGGAACGACACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAG 632
        |||||:::||||||   ||||||||||||||||||||||||||||||||
      245 rgArgLysAspThrArgThrMetThrArgValLeuTyrGluArgAlaGln 261

    631 GCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGG 582
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      262 AlaGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGl 278

    581 GCTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGC 532
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      278 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeuL 295

    531 TCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCC 482
        ||||||||||||||||||||||||||||||||||||||   ||||||
      295 euAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 311

    481 GTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTA 432
        :::|||||||||||||||||||||||||||:::||||||||||||||||||
      312 IleSerArgPheValGluGluAlaLeuArgHisHisProProValProTy 328

    431 CACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGC 382
        ||||||||||||||||||||||||||||||:::|||   ||||||||||
      328 rThrLeuTrpArgPheAlaAlaThrGluValValIleArgGlyValArgL 345

    381 TGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGAC 332
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      345 euProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAsp 361

    331 GGCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTG 282
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      362 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerAr 378

    281 GCGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGC 232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      378 gArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnL 395

    231 TCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTC 182
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      395 euAlaGlnLeuGluSerArgThrMetIleGlyValLeuArgSerArgPhe 411

    181 CCCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAA 132
        |||:::|||||||||||||||||||||:::|||||||||||||||||||
      412 ProGlnAlaArgLeuAlaValProTyrGluGluLeuArgTrpCysArgLy 428

    131 GGGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGC 86
        ||||||||||||||||||||||||:::|||||||||||||||||
      428 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValTrpLeuArg 443

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-653-650A-9

seq_documentation_block:
; Sequence 9, Application US/08653650A
; Patent No. 5976830
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/653,650A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Golrick, Mary E.
;       REGISTRATION NUMBER:  34829
;       REFERENCE/DOCKET NUMBER:  22727/00131
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  216-622-8458
;       TELEFAX:  216-241-0816
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
        LENGTH:  443 amino acids
        TYPE:  amino acid
        STRANDEDNESS:  single
        TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-653-650A-9


alignment_scores:
            Quality: 2104.50              Length:    432
              Ratio:  4.987                Gaps:      1
 Percent Similarity:  97.685   Percent Identity: 93.519

alignment_block:
 US-09-673-254-1/rev x US-08-653-650A-9   ..

 Align seg 1/1  to: US-08-653-650A-9  from: 1  to: 443

    1378 GTACCCCGCGGCGGCGGGCGG...TGCCTCGTGAGCGGCGAGGCGCCCCG 1332
         ||||||   |||:::::::::   ||||||:::||||||||||||||||
      12 ValProTyrGlyAsnSerGluGlyCysLeuMetSerGlyGluAlaProAr 28

    1331 GGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGC 1282
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
      28 gValAlaValAspProPheSerCysProMetMetThrMetGlnArgLysP 45

    1281 CCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAAC 1232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      45 roGluValHisAspAlaPheArgGluAlaGlyProValValGluValAsn 61

    1231 GCCCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCG 1182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      62 AlaProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaAr 78

    1181 CGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCG 1132
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
      78 gGluValLeuAlaAspProArgPheValLysAspProAspLeuAlaProT 95

    1131 CCGCCTGGCGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTG 1082
         ::||||||||||||||||||||||||||||||||||||||||||||||||
      95 hrAlaTrpArgGlyValAspAspGlyLeuAspIleProValProGluLeu 111

    1081 CGTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCG 1032
         ||||||||||||||||||||||||||||||||   ||||||||||||||
      112 ArgProPheThrLeuIleAlaValAspGlyGluAspHisArgArgLeuAr 128

    1031 CCGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGG 982
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      128 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrA 145

      981 ATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCC 932
          ||||||||||||||||||   |||||||||||||||||||||||:::
      145 spArgIleAlaAlaIleAlaAspArgLeuLeuThrGluLeuAlaAspSer 161

      931 TCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCA 882
          |||   ||||||||:::|||||||||||||||||||||||||||||||||
      162 SerAspArgSerGlyGluProAlaGluLeuIleGlyGlyPheAlaTyrHi 178

      881 CTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATC 832
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      178 sPheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspP 195

      831 CGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGC 782
          ||||||||||||||||||||:::||||||||||||||||||||||||||
      195 roAlaMetAlaArgGluAlaValGlyValLeuLysAlaLeuGlyLeuGly 211

      781 GGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCC 732
          ||||||||||||:::|||||||||||||||||||||||||   |||||
      212 GlyProGlnSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 228

      731 GGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCC 682
          ||||||||||||||||||||||||||||||||||||||||||:::||||
      228 oAspThrSerAlaLeuGluSerLeuLeuLeuGluAlaValHisAlaAlaA 245

      681 GGCGGAACGACACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAG 632
          ||||||:::||||||   ||||||||||||||||||||||||||||||||
      245 rgArgLysAspThrArgThrMetThrArgValLeuTyrGluArgAlaGln 261

      631 GCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGG 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      262 AlaGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGl 278

      581 GCTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGC 532
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      278 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeuL 295

      531 TCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCC 482
          ||||||||||||||||||||||||||||||||||||||   ||||||
      295 euAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 311

      481 GTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTA 432
          :::|||||||||||||||||||||||||||:::||||||||||||||||
      312 IleSerArgPheValGluGluAlaLeuArgHisHisProProValProTy 328

      431 CACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGC 382
          ||||||||||||||||||||||||||||||:::|||   ||||||||||
      328 rThrLeuTrpArgPheAlaAlaThrGluValValIleArgGlyValArgL 345

      381 TGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGAC 332
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      345 euProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAsp 361

      331 GGCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTG 282
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      362 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerAr 378

      281 GCGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGC 232
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      378 gArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnL 395

      231 TCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTC 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      395 euAlaGlnLeuGluSerArgThrMetIleGlyValLeuArgSerArgPhe 411

      181 CCCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAA 132
          |||:::|||||||||||||||||||:::|||||||||||||||||||||
      412 ProGlnAlaArgLeuAlaValProTyrGluGluLeuArgTrpCysArgLy 428

      131 GGGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGC 86
          ||||||||||||||||||||||||:::|||||||||||||||||||
      428 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValTrpLeuArg 443

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-982-8

seq_documentation_block:
; Sequence 8, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
```

```
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/096,982
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-096-982-8

alignment_scores:
            Quality: 2093.50              Length:    432
              Ratio:  4.973                Gaps:      1
 Percent Similarity:  97.454   Percent Identity: 93.056

alignment_block:
 US-09-673-254-1/rev x US-09-096-982-8   ..

 Align seg 1/1  to: US-09-096-982-8  from: 1  to: 474

    1378 GTACCCCGCGGCGGCGGGCGG...TGCCTCGTGAGCGGCGAGGCGCCCCG 1332
         ||||||   ||:::::::::   ||||||:::||||||||||||||||
      43 ValProTyrGlyAsnSerGluGlyCysLeuMetSerGlyGluAlaProAr 59

    1331 GGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGC 1282
         ||||||||||||||||||:::||||||||||||||||||||||||||||
      59 gValAlaValAspProPheSerCysProMetMetThrMetGlnArgLysP 76

    1281 CCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAAC 1232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      76 roGluValHisAspAlaPheArgGluAlaGlyProValValGluValAsn 92

    1231 GCCCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCG 1182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      93 AlaProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaAr 109

    1181 CGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCG 1132
         ||||||||||||||||||||||||||||||   ||||||||||||||:
     109 gGluValLeuAlaAspProArgPheValLysGlyProAspLeuAlaProT 126

    1131 CCGCCTGGCGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTG 1082
         ::||||||||||||||||||||||||||||||||||||||||||||||||
     126 hrAlaTrpArgGlyValAspAspGlyLeuAspIleProValProGluLeu 142

    1081 CGTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCG 1032
         ||||||||||||||||||||||||||||||   ||||||||||||||
     143 ArgProPheThrLeuIleAlaValAspGlyGluAspHisArgArgLeuAr 159

    1031 CCGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGG 982
         ||||||||||||||||||||||||||||||||||||||||||||||||||

    159 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrA 176

    981 ATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCC 932
        ||||||||||||||||||   |||||||||||||||||||||||:::
    176 spArgIleAlaAlaIleAlaAspArgLeuLeuThrGluLeuAlaAspSer 192

    931 TCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCA 882
        |||   ||||||||:::||||||||||||||||||||||||||||||||
    193 SerAspArgSerGlyGluProAlaGluLeuIleGlyGlyPheAlaTyrHi 209

    881 CTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATC 832
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    209 sPheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspP 226

    831 CGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGC 782
        ||||||||||||||||||||||:::||||||||||||||||||||||||
    226 roAlaMetAlaArgGluAlaValGlyValLeuLysAlaLeuGlyLeuGly 242

    781 GGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCC 732
        ||||||||||||:::|||||||||||||||||||||||||||   |||||
    243 GlyProGlnSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 259

    731 GGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCC 682
        ||||||||||||||||||||||||||||||||||||||||:::||||
    259 oAspThrSerAlaLeuGluSerLeuLeuLeuGluAlaValHisAlaAlaA 276

    681 GGCGGAACGACACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAG 632
        |||||:::|||||||   ||||||||||||||||||||||||||||||
    276 rgArgLysAspThrArgThrMetThrArgValLeuTyrGluArgAlaGln 292

    631 GCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGG 582
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    293 AlaGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGl 309

    581 GCTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGC 532
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    309 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeuL 326

    531 TCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCC 482
        ||||||||||||||||||||||||||||||||||||||   ||||||
    326 euAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 342

    481 GTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTA 432
        :::||||||||||||||||||||||||||||:::||||||||||||||||
    343 IleSerArgPheValGluGluAlaLeuArgHisHisProProValProTy 359

    431 CACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGC 382
        |:::|||||||||||||||||||||||||||:::|||   |||||||||
    359 rSerLeuTrpArgPheAlaAlaThrGluValValIleArgGlyValArgL 376

    381 TGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGAC 332
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    376 euProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAsp 392

    331 GGCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTG 282
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    393 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerAr 409

    281 GCGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGC 232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    409 gArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnL 426

    231 TCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTC 182
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    426 euAlaGlnLeuGluSerArgThrMetIleGlyValLeuArgSerArgPhe 442

    181 CCCGAGGCCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAA 132
        |||:::||||||||||||||||||:::||||||||||||||||||||||
    443 ProGlnAlaArgLeuAlaValProTyrGluGluLeuArgTrpCysArgLy 459

    131 GGGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGC 86
        ||||||||||||||||||||||:::||||||||||||||||||
    459 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValTrpLeuArg 474
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-653-650A-8

seq_documentation_block:
; Sequence 8, Application US/08653650A
; Patent No. 5976830
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
       STREET:  800 Superior Avenue, Suite 1400
       CITY:  Cleveland
       STATE:  Ohio
       COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/653,650A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-653-650A-8

alignment_scores:
           Quality: 2093.50              Length:    432
             Ratio:  4.973                Gaps:      1
[redacted]cent Similarity:  97.454   Percent Identity: 93.056

alignment_block:
 US-09-673-254-1/rev x US-08-653-650A-8   ..

 Align seg 1/1  to: US-08-653-650A-8  from: 1  to: 474

    1378 GTACCCCGCGGCGGCGGGCGG...TGCCTCGTGAGCGGCGAGGCGCCCCG 1332
         ||||||   |||:::::::::   ||||||:::||||||||||||||||
      43 ValProTyrGlyAsnSerGluGlyCysLeuMetSerGlyGluAlaProAr 59

    1331 GGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGC 1282
         ||||||||||||||||||:::||||||||||||||||||||||||||||
      59 gValAlaValAspProPheSerCysProMetMetThrMetGlnArgLysP 76

    1281 CCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAAC 1232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      76 roGluValHisAspAlaPheArgGluAlaGlyProValValGluValAsn 92

    1231 GCCCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCG 1182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      93 AlaProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaAr 109

    1181 CGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCG 1132
         |||||||||||||||||||||||||||||   ||||||||||||||:
      109 gGluValLeuAlaAspProArgPheValLysGlyProAspLeuAlaProT 126

    1131 CCGCCTGGCGGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTG 1082
         ::||||||||||||||||||||||||||||||||||||||||||||||||
     126 hrAlaTrpArgGlyValAspAspGlyLeuAspIleProValProGluLeu 142

    1081 CGTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCG 1032
         ||||||||||||||||||||||||||||||   ||||||||||||||
     143 ArgProPheThrLeuIleAlaValAspGlyGluAspHisArgArgLeuAr 159

    1031 CCGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGG 982
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     159 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrA 176

     981 ATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCC 932
         ||||||||||||||||||   |||||||||||||||||||||:::
     176 spArgIleAlaAlaIleAlaAspArgLeuLeuThrGluLeuAlaAspSer 192

     931 TCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCA 882
         |||   |||||||||:::||||||||||||||||||||||||||||||||
     193 SerAspArgSerGlyGluProAlaGluLeuIleGlyGlyPheAlaTyrHi 209

     881 CTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATC 832
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     209 sPheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspP 226

     831 CGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGC 782
         ||||||||||||||||||||:::||||||||||||||||||||||||||
     226 roAlaMetAlaArgGluAlaValGlyValLeuLysAlaLeuGlyLeuGly 242

     781 GGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCCGGGGGCGTGCC 732
         ||||||||||||:::|||||||||||||||||||||||||||   |||||
     243 GlyProGlnSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 259

     731 GGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCC 682
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
     259 oAspThrSerAlaLeuGluSerLeuLeuLeuGluAlaValHisAlaAlaA 276

     681 GGCGGAACGACACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAG 632
         |||||:::||||||   ||||||||||||||||||||||||||||||||
     276 rgArgLysAspThrArgThrMetThrArgValLeuTyrGluArgAlaGln 292

     631 GCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGG 582
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     293 AlaGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGl 309

     581 GCTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGC 532
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     309 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeuL 326

     531 TCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCC 482
         ||||||||||||||||||||||||||||||||||||   ||||||
     326 euAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 342

     481 GTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTA 432
         :::||||||||||||||||||||||||||:::||||||||||||||||||
     343 IleSerArgPheValGluGluAlaLeuArgHisHisProProValProTy 359

     431 CACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGC 382
         |:::|||||||||||||||||||||||||:::|||   ||||||||||
     359 rSerLeuTrpArgPheAlaAlaThrGluValValIleArgGlyValArgL 376

     381 TGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGAC 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     376 euProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAsp 392

     331 GGCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTG 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     393 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerAr 409

     281 GCGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGC 232
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     409 gArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnL 426
```

```
     231 TCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTC 182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     426 euAlaGlnLeuGluSerArgThrMetIleGlyValLeuArgSerArgPhe 442

     181 CCCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAA 132
         |||:::|||||||||||||||||||||:::||||||||||||||||||||
     443 ProGlnAlaArgLeuAlaValProTyrGluGluLeuArgTrpCysArgLy 459

     131 GGGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGC 86
         |||||||||||||||||||||||:::|||||||||||||||||||
     459 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValTrpLeuArg 474

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-096-982-5

seq_documentation_block:
; Sequence 5, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/096,982
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    -096-982-5


alignment_scores:
            Quality: 2076.00              Length:    422
              Ratio:  5.027                Gaps:      0
 Percent Similarity:  97.867   Percent Identity: 94.076

alignment_block:
 US-09-673-254-1/rev x US-09-096-982-5   ..

 Align seg 1/1  to: US-09-096-982-5  from: 1  to: 422

    1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
         :::|||||||||||||||||||||||||||||||||||||||:::|||||||
       1 MetSerGlyGluAlaProArgValAlaValAspProPheSerCysProMe 17

    1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

    1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50

    1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValLy 67

    1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCG 1102
         |   |||||||||||||:::|||||||||||||||||||||||||||||||
      67 sGlyProAspLeuAlaProThrAlaTrpArgGlyValAspAspGlyLeuA 84

    1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

    1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
         |||   ||||||||||||||||||||||||||||||||||||||||||||
     101 GluAspHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117

    1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
         |||||||||||||||||||||||||||||||||||||   |||||||
     117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaAspArgLeuL 134

     951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
         ||||||||||||||:::|||   |||||||||:::|||||||||||||
     134 euThrGluLeuAlaAspSerSerAspArgSerGlyGluProAlaGluLeu 150

     901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 IleGlyGlyPheAlaTyrHisPheProLeuLeuValIleCysGluLeuLe 167

     851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
         ||||||||||||||||||||||||||||||||||||||||||:::||||
     167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValGlyValL 184

     801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
         ||||||||||||||||||||||||||||||||||:::|||||||||||||
     184 euLysAlaLeuGlyLeuGlyGlyProGlnSerAlaGlyGlyAspGlyThr 200

     751 GACCCTGCCGGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
         ||||||||||||   ||||||||||||||||||||||||||||||||||
     201 AspProAlaGlyAspValProAspThrSerAlaLeuGluSerLeuLeuLe 217

     701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
         ||||||||||||:::|||||||:::||||||   ||||||||||||||
     217 uGluAlaValHisAlaAlaArgArgLysAspThrArgThrMetThrArgV 234

     651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250

     601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrGl 267

     551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284

     501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
         ||||||||||   ||||||:::|||||||||||||||||||||||||||
     284 laAspAlaAspGlyAspAlaIleSerArgPheValGluGluAlaLeuArg 300

     451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
         :::|||||||||||||||||:::|||||||||||||||||||||||||||
     301 HisHisProProValProTyrSerLeuTrpArgPheAlaAlaThrGluVa 317

     401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
         |:::|||   ||||||||||||||||||||||||||||||||||||||||
     317 lValIleArgGlyValArgLeuProArgGlyAlaProValLeuValAspI 334

     351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||
334 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350

301 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
    ||||||||||||||||||  ||||||||||||||||||||||||||||||
351 HisProAspArgProSerArgArgArgLeuThrPheGlyAspGlyProHi 367

251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
    ||||||||||||||||||||||||||||||||||||||||||||||||||
367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384

201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
    ||||||||||||||||||||||:::||||||||||||||||||||||:::
384 lyValLeuArgSerArgPheProGlnAlaArgLeuAlaValProTyrGlu 400

151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
    ||||||||||||||||||||||||||||||||||||||||||||||:::||
401 GluLeuArgTrpCysArgLysGlyAlaGlnThrAlaArgLeuThrAspLe 417

101 GCCCGTCTGGCTGCGC 86
    ||||||||||||||||
417 uProValTrpLeuArg 422
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 17:58:58 ; Search time 3477.18 Seconds
(without alignments)
17272.376 Million cell updates/sec

Title: US-09-673-254-1
Perfect score: 2870
Sequence: 1 ggatccgcaccgggtacacg.....catcgatgcgcgggtaccgg 2870

[redacted]g table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2870	100.0	6985	1	SPU77891	U77891 Streptomyce
c 2	2567.8	89.5	3196	1	SSU50973	U50973 Streptomyce
c 3	2542.2	88.6	3196	6	AR078146	AR078146 Sequence
c 4	2542.2	88.6	3196	6	AR083230	AR083230 Sequence
c 5	1732.4	60.4	1932	1	AF403708	AF403708 Streptomy
c 6	1269	44.2	1269	6	I84584	I84584 Sequence 1
c 7	1221.6	42.6	2081	6	AR078148	AR078148 Sequence
c 8	1221.6	42.6	2081	6	AR083232	AR083232 Sequence
c 9	1102.2	38.4	3013	6	AR078147	AR078147 Sequence
c 10	1102.2	38.4	3013	6	AR083231	AR083231 Sequence
c 11	1046	36.4	1569	6	AR075142	AR075142 Sequence
c 12	1044.4	36.4	1569	6	AR144737	AR144737 Sequence
c 13	592	20.6	864	6	AR075141	AR075141 Sequence
c 14	592	20.6	864	6	AR144736	AR144736 Sequence
15	576.4	20.1	5985	1	SCU43704	U43704 Streptomyce
16	192.6	6.7	33320	1	SC9B10	AL009204 Streptomy
17	185	6.4	36028	1	SCE41	AL442120 Streptomy
c 18	181.2	6.3	123580	1	AF263912	AF263912 Streptomy
c 19	181.2	6.3	125401	6	AX211739	AX211739 Sequence
20	176	6.1	41807	1	SC6D11	AL158061 Streptomy
c 21	173.8	6.1	33810	1	SCE6	AL353832 Streptomy
22	160.6	5.6	30941	1	SC4C6	AL079355 Streptomy
c 23	154.8	5.4	7870	1	STE18574	Y18574 Streptomyce
c 24	152.6	5.3	63734	1	AF127374	AF127374 Streptomy
c 25	149.2	5.2	36583	1	SC5H1	AL049863 Streptomy
c 26	148.2	5.2	14750	1	SCF1	AL117322 Streptomy
c 27	147.4	5.1	104326	1	AB070940	AB070940 Streptomy
c 28	144.6	5.0	113193	1	AF357202	AF357202 Streptomy
29	144.4	5.0	46166	1	SC8F11	AL353864 Streptomy
c 30	142	4.9	65140	6	AX211705	AX211705 Sequence
c 31	141.8	4.9	24268	1	AB070947	AB070947 Streptomy
c 32	140.2	4.9	3912	1	REQ242746	AJ242746 Rhodococc
c 33	139.8	4.9	1257	6	AX097457	AX097457 Sequence
c 34	139.8	4.9	14759	1	AE004755	AE004755 Pseudomon
c 35	138.6	4.8	20535	1	SC2G4	AL360034 Streptomy
36	137.8	4.8	37750	1	SC6D7	AL133213 Streptomy
c 37	137.4	4.8	1470	1	AF087022	AF087022 Streptomy
c 38	137.4	4.8	4257	6	AR144708	AR144708 Sequence
c 39	137.4	4.8	4257	6	AR145616	AR145616 Sequence
c 40	137.4	4.8	4342	1	AF079139	AF079139 Streptomy
c 41	137.4	4.8	6633	14	HEHSV1G3	X06461 Herpes simp
42	137.4	4.8	26245	14	HS1US	L00036 Human herpe
c 43	137.4	4.8	26245	14	HS1US	L00036 Human herpe
44	137.4	4.8	152261	14	HE1CG	X14112 Herpes simp
c 45	137.4	4.8	152261	14	HE1CG	X14112 Herpes simp

ALIGNMENTS

RESULT 1
SPU77891
LOCUS SPU77891 6985 bp DNA linear BCT 13-JAN-1999
DEFINITION Streptomyces peucetius daunorubicin C-14 hydroxylase (doxA), doxorubicin biosynthesis enzyme DnrV (dnrV), daunorubicin C-13 ketoreductase (dnrU), daunorubicin acyl carrier protein (dpsG), daunorubicin biosynthesis enzyme (dpsH), doxorubicin biosynthesis protein DnmT (dnmT), and putative baumycin biosynthesis protein (dnrH) genes, complete cds.
ACCESSION U77891
VERSION U77891.1 GI:3778994
KEYWORDS .
SOURCE Streptomyces peucetius.
ORGANISM Streptomyces peucetius
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 3983 to 6985)
AUTHORS Scotti,C. and Hutchinson,C.R.
TITLE Enhanced antibiotic production by manipulation of the Streptomyces peucetius dnrH and dnmT genes involved in doxorubicin (adriamycin) biosynthesis
JOURNAL J. Bacteriol. 178 (24), 7316-7321 (1996)
MEDLINE 97113559

REFERENCE 2 (bases 1 to 6985)
AUTHORS Lomovskaya,N., Otten,S.L., Doi-Katayama,Y., Fonstein,L., Liu,X.-C., Takatsu,T., Inventi,A., Filippi,S., Torti,F., Colombo,A.L. and Hutchinson,C.R.
TITLE Doxorubicin overproduction in Streptomyces peucetius: cloning and characterization of the dnrU ketoreductase and dnrV genes and the doxA cytochrome P-450 hydroxylase gene
JOURNAL J. Bacteriol. 181 (1), 305-318 (1999)
MEDLINE 99084971
REFERENCE 3 (bases 3893 to 6985)
AUTHORS Scotti,C. and Hutchinson,C.R.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1996) School of Pharmacy, University of Wisconsin, 425 N. Charter St., Madison, WI 53706, USA
REFERENCE 4 (bases 1 to 6985)
AUTHORS Lomovskaya,N., Otten,S.L., Doi-Katayama,Y., Fonstein,L., Liu,X.-C., Takatsu,T., Inventi,A., Filippi,S., Torti,F., Colombo,A.L. and Hutchinson,C.R.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1998) School of Pharmacy, University of Wisconsin, 425 N. Charter St., Madison, WI 53706, USA
[redacted]MARK Sequence update by submitter
[redacted]ENT On Oct 22, 1998 this sequence version replaced gi:1684919.
[redacted]TURES Location/Qualifiers
source 1. .6985
/organism="Streptomyces peucetius"
/strain="ATCC 29050"
/db_xref="ATCC:29050"
/db_xref="taxon:1950"
gene complement(83. .1330)
/gene="doxA"
CDS complement(83. .1330)
/gene="doxA"
/note="CYP450 hydroxylase"
/codon_start=1
/transl_table=11
/product="daunorubicin C-14 hydroxylase"
/protein_id="AAD04715.1"
/db_xref="GI:3778995"
/translation="MAVDPFACPMMTMQRKPEVHDAFREAGPVVEVNAPAGGPAWVIT
DDALAREVLADPRFVKDPDLAPAAWRGVDDGLDIPVPELRPFTLIAVDGEAHRRLRRI
HAPAFNPRRLAERTDRIAAIAGRLLTELADASGRSGKPAELIGGFAYHFPLLVICELL
GVPVTDPAMAREAVSVLKALGLGGPQSGGGDGTDPAGGVPDTSALESLLLEAVHSARR
NDTPTMTRVLYERAQAEFGSVSDDQLVYMITGLIFAGHDTTGSFLGFLLAEVLAGRLA
ADADEDAVSRFVEEALRYHPPVPYTLWRFAATEVTIGGVRLPRGAPVLVDIEGTNTDG
RHHDAPHAFHPDRPSWRRLTFGDGPHYCIGEQLAQLESRTMIGVLRSRFPEARLAVPY
DELRWCRKGAQTARLTELPVWLR"
gene complement(1348. .2175)
/gene="dnrV"
CDS complement(1348. .2175)
/gene="dnrV"
/codon_start=1
/transl_table=11
/product="doxorubicin biosynthesis enzyme DnrV"
/protein_id="AAD04716.1"
/db_xref="GI:3778996"
/translation="MTRFAPGAPAWFDLGSPDVAASADFYTGLFGWTATVVSDPGAGG
YTTFSSDGKLVAAVARHQIDTPYHRPYGPGNDQHGMPAIWTVYFATDDADALTKRVET
AGGEVIMTPMDVLGLGRMAVFADPAGAAFAVWRKGVMEGAEVTGVPGSVGWVELVTDG
IGAARDFYPATLGLAPADTGLKGVTDPVWHIGDTPVAGTQELGVTGAVRPHWAVLFAV
HDCDATVRRAVELGGSVENERADTPRGRRADLLDPHGAGFSVVELREGYPAAAGGAS"
gene complement(2279. .3142)
/gene="dnrU"
CDS complement(2279. .3142)
/gene="dnrU"
/codon_start=1
/transl_table=11
/product="daunorubicin C-13 ketoreductase"
/protein_id="AAD04717.1"
/db_xref="GI:3778997"
/translation="MTASTPHHGTPRGGLSGRTVLVTGATSGIGRAAALAVARQGARV
VLVGRDPERLRTVTNEVARTAGPAPDAFRADFAELRQVRDLGERLRDRYPRIDVMASN
AGGMFWSRTTTQDGFEATIQVNHLAGFLLARLLRERLAGGRLILTSSDAYTQGRIDPD
DLNGDRHRYSAGQAYGTSKQANIMTAAEAARRWPDVLAVSYHPGEVRTRIGRGTVASS
YFRFNPFLRSAAKGADTLVWLASAPAEELTTGGYYSDRRLSPVSGPTADAGLAAKLWE
AGAAAVGDTAH"
gene complement(3139. .3393)
/gene="dpsG"
CDS complement(3139. .3393)
/gene="dpsG"
/note="component of daunorubicin/doxorubicin polyketide synthase"
/codon_start=1
/transl_table=11
/product="daunorubicin acyl carrier protein"
/protein_id="AAD04718.1"
/db_xref="GI:3778998"
/translation="MAELSLAELREIMRQSLGEDEVPDLADADTVTFEDLGLDSLAVL
ETVNHIERTYGVKLPEEELAEVRTPHSMLIFVNERLRAAA"
gene 3493. .4077
/gene="dpsH"
CDS 3493. .4077
/gene="dpsH"
/codon_start=1
/transl_table=11
/product="daunorubicin biosynthesis enzyme"
/protein_id="AAD04719.1"
/db_xref="GI:3778999"
/translation="MDRRLSALPQDARNRLGTAPAGRRTRPGQDRAARGHHTTQGVTM
DALPADDAPTTESADCTDPGIYLQVQRFYARQMQLLDDGRAEDWARTFAQDGVFEVGG
EAVRGAADIARAARRTTDRFAAEGITRRHWIGMLTVDGKPDEVTARSYAVVLETPRDG
DPVLRRSTVCTDILVREAGEWRVKHRTVTRDGLD"
gene 4129. .5646
/gene="dnmT"
CDS 4129. .5646
/gene="dnmT"
/note="catalyzes unknown step in biosynthesis of daunosamine, an intermediate of doxorubicin biosynthesis; deoxysugar biosynthesis enzyme"
/codon_start=1
/transl_table=11
/product="doxorubicin biosynthesis protein DnmT"
/protein_id="AAD04713.1"
/db_xref="GI:1684920"
/translation="MTAQIARSVLARDGLGSGMDRFWAWYADRSAQVVHRTERIPLDD
LKGWSRHPVTGTVSHHTGRFFSIEGLDVHIPGAPVPRWSQPIVNQPEVGILGFLVKER
HGVLHCLVQAKFEPGNPGGLQLSPTVQATRSNYTRVHGGKAIPYLEHFRDTAERQVVA
DVLQSEQGSWFYRKRNRNMIVQVIDEVPLHEDFHWLTLGQLHRLLGVENLVNMDARTV
LACLPFASEGAHPLADAEPAGADGPGGGHPVVPARPDGGEGRVDTGFHRSLVRSCAAA
EGSLHSTVDIVSWIADLRSRTDVVTRPAALNALPHWYERDGAIAHESGRFLEVMAVDV
TAASREVPGWSQPMIEPKDQGVAAFLVRRIDGVLHVLAHARVEPGYVDVVEIAPTVQC
TPGSLHALPAEARPRFLDAVLEAPPERVRYATVLAEEGGRFYHAVNTYMIVEADHDIP
DGGEYRWLTLHQLVGLLRHSHYVNVQARTLVACLHSLSVGSPVTRSAVPSAPPRR"
gene 5654. .6985
/gene="dnrH"
CDS 5654. .6985
/gene="dnrH"
/note="catalyzes formation of baumycins, C-4' glycosides of doxorubicin; putative glycosyl transferase; DnrH"
/codon_start=1
/transl_table=11
/product="putative baumycin biosynthesis protein"
/protein_id="AAD04714.1"
/db_xref="GI:1684921"
/translation="MRVLFATMAARSHVYAQVTLASALRTAGHEVLVASQPDVLDDIV
RAGLTAVRIGEDLNIEEETREANASFEDDRNLGGLAMSNSRDDPFPWDHALGMFTAMT
AMVFQNVCPEPMVDDLVGLARDWRPDLVVWDPLTLAGPVAARLSGAAHARLLFGPDQM
GRNRTAFRALLDRQPPELRDDPLAEWLTWTLERCGGSAGDMSEELVLGQWTIDPTPPS
MRIPLDLPCVPVRYVPYNGPSLLPGWLREPPRHPRRLCLTLGVSLGEATGAGTVAASD
VLAAVDGLDVEVVATLPRELREELGTLPANVRAVDFVPLNALLPSCSGIIHHGGSGTF
MTALVHAIPQLIVPDMMWDAMEKAHGLARSGAGSYVDARDVSPELLRERVLALLDDPS
YAAGARRVRAEIVGTPSPNDIVPVLERLTAEHRAGGAEGGPALKSPSTGGA"
BASE COUNT 958 a 2535 c 2543 g 949 t
ORIGIN

```
Query Match             100.0%;  Score 2870;  DB 1;  Length 6985;
Best Local Similarity   100.0%;  Pred. No. 6.7e-230;
Matches 2870;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGATCCGCACCGGGTACACGGCACGGGACCGCCCACCGCGCGGTGCGCGGTGGGCGGTCC 60

Qy      61 cgtgccggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGTGCCGGTCGCGGCCGGCGGATCAGCGCAGCCAGACGGGCAGTTCGGTGAGCCGCGCCG 120

Qy     121 tctgggccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       121 TCTGGGCCCCTTCCGGCACCACCGCAACTCGTCGTACGGCACGGCCAGTCGGGCCTCGG 180

       181 ggaacctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGAACCTGCTGCGCAGTACGCCGATCATCGTGCGCGACTCCAGCTGGGCGAGCTGCTCCC 240

Qy     241 cgatgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggt 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGATGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCACGAGGGACGGTCCGGGT 300

Qy     301 ggaaggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccacca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGAAGGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCA 360

Qy     361 gcaccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCACCGGCGCTCCGCGGGGCAGCCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACC 420

Qy     421 tccacaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggaga 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCCACAACGTGTAGGGCACCGGCGGGTGGTAGCGCAGCGCCTCCTCCACGAACCGGGAGA 480

Qy     481 cggcgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CGGCGTCCTCGTCGGCATCCGCCGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGC 540

Qy     541 ccaggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacga 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCAGGAAGGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACGA 600

Qy     601 gctggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtca 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       601 GCTGGTCGTCGGAGACCGAGCCGAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCA 660

       661 tggtcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccaggg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TGGTCGGGGTGTCGTTCCGCCGGGCTGAGTGCACGGCTTCGAGGAGCAGGCTCTCCAGGG 720

Qy     721 ccgaggtgtccggcacgcccccggcagggtccgtgccgtcaccccgccgctctgcgggc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCGAGGTGTCCGGCACGCCCCCGGCAGGGTCCGTGCCGTCACCCCCGCCGCTCTGCGGGC 780

Qy     781 cgccgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtga 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CGCCGAGGCCGAGTGCCTTGAGAACGCTGACGGCCTCGCGGGCCATCGCCGGATCGGTGA 840

Qy     841 ccggcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccga 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCGGCACACCGAGCAGCTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGA 900

Qy     901 tcagctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCAGCTCGGCCGGTTTGCCCGACCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGC 960

Qy     961 cggcgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CGGCGATCGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTG 1020

Qy    1021 cgtggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggac 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGTGGATGCGGCGCAGGCGCCGGTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGAC 1080

Qy    1081 gcagctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcga 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GCAGCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGCGGGGGCGA 1140

Qy    1141 ggtcggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcgg 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GGTCGGGGTCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGG 1200

Qy    1201 tgatgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctccc 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TGATGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACCGGGCCCGCCTCCC 1260

Qy    1261 ggaaggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggt 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GGAAGGCGTCGTGCACCTCGGGCTTGCGCTGCATGGTCATCATGGGACACGCGAACGGGT 1320

Qy    1321 cgacggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtaccc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CGACGGCCACCCGGGGCGCCTCGCCGCTCACGAGGCACCGCCCGCCGCCGCGGGGTACCC 1380

Qy    1381 ctcccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccg 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CTCCCGCAGTTCGACCACCGAGAAGCCGGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCG 1440

Qy    1441 cccccctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccg 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CCCCCTGGGCGTGTCGGCGGGCTCGTTCTCGACGGAGCCGCCGAGTTCAACGGCGCGCCG 1500

Qy    1501 gaccgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgcc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GACCGTCGCGTCGCAGTCGTGCACGGCGAACAGCACGGCCCAGTGCGGCCGTACCGCGCC 1560

Qy    1561 ggtgacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtc 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GGTGACGCCCAGCTCCTGGGTGCCGGCGACCGGTGTGTCACCGATGTGCCAGACCGGGTC 1620

Qy    1621 ggtgacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GGTGACGCCCTTCAGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCCGGGTAGAAGTC 1680

Qy    1681 ccgggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcc 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCGGGCGGCCCCGATGCCGTCGGTCACCAGCTCGACCCAGCCGACCGAGCCGGGCACGCC 1740

Qy    1741 cgtcacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggg 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CGTCACCTCCGCGCCCTCCATGACTCCCTTGCGCCAGACCGCGAACGCGGCCCCGGCGGG 1800

Qy    1801 gtcggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctc 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 GTCGGCGAAGACCGCCATCCGGCCGAGGCCGAGGACGTCCATCGGAGTCATGATGACCTC 1860

Qy    1861 gccgcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacac 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 GCCGCCCGCCGTCTCGACCCGCTTGGTCAGTGCGTCGGCGTCGTCGGTGGCGAAGTACAC 1920

Qy    1921 ggtccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggg 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GGTCCAGATGGCCGGCATGCCGTGCTGGTCGTTCCCGGGCCCGTACGGCCGGTGGTAGGG 1980

Qy    1981 ggtgtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgt 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 GGTGTCGATCTGGTGGCGGGCGACCGCGGCGACCAGCTTCCCGTCGGAGCTGAACGTCGT 2040

Qy    2041 gtatccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgta 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 GTATCCCCCGGCGCCCGGGTCGCTGACCACGGTGGCGGTCCAGCCGAACAGGCCGGTGTA 2100

Qy    2101 gaagtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccggg 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    2101 GAAGTCGGCCGAGGCGGCGACATCGGGCGAACCGAGGTCGAACCATGCGGGGGCGCCGGG 2160

Qy    2161 cgcgaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgc 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CGCGAACCTGGTCACGAATCGTTCCTTTCGATGGATCGGCACACGAGCGTCTGCGCTCGC 2220

Qy    2221 ggatgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtc 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 GGATGAGACGGACATCTCGCGGATGAGACGGACATGCGGGCGGGGCGGGCCGCCGCCGTC 2280

Qy    2281 agtgcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccgg 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 AGTGCGCGGTGTCGCCGACGGCGGCCGCGCCGGCCTCCCAGAGCTTCGCCGCGAGGCCGG 2340

Qy    2341 cgtcggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtgg 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CGTCGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGG 2400

Qy    2401 tcaactcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcgg 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 TCAACTCCTCGGCCGGCGCGGACGCCAGCCACACGAGGGTGTCGGCGCCCTTCGCCGCGG 2460

      2461 agcgcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcggg 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 AGCGCAGGAAGGGGTTGAACCGGAAGTAGGACGAGGCGACCGTGCCCCGTCCGATGCGGG 2520

Qy    2521 tgcggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcct 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TGCGGACCTCACCGGGGTGATAGCTGACCGCCAGCACGTCCGGCCAGCGCCTGGCGGCCT 2580

Qy    2581 ccgccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagc 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CCGCCGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCGGCGCTGTAGC 2640

Qy    2641 ggtgacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacg 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 GGTGACGGTCGCCGTTGAGGTCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACG 2700

Qy    2701 aggtgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagc 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 AGGTGAGGATCAGCCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGC 2760

Qy    2761 ctgcgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgacc 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 CTGCGAGGTGATTGACCTGGATGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACC 2820

Qy    2821 agaacatgccgccggcgttgctggccatgacatcgatgcgcgggtaccgg 2870
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 AGAACATGCCGCCGGCGTTGCTGGCCATGACATCGATGCGCGGGTACCGG 2870
```

[redacted]LT 2
973/c

LOCUS SSU50973 3196 bp DNA linear BCT 14-SEP-1996
DEFINITION Streptomyces sp. daunomycin C-14 hydroxylase (doxA), gene, complete cds, and (dauI) gene, partial cds.
ACCESSION U50973
VERSION U50973.1 GI:1354266
KEYWORDS .
SOURCE Streptomyces sp.
ORGANISM Streptomyces sp.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 3196)
AUTHORS Dickens,M.L. and Strohl,W.R.
TITLE Isolation and characterization of a gene from Streptomyces sp. strain C5 that confers the ability to convert daunomycin to doxorubicin on Streptomyces lividans TK24
JOURNAL J. Bacteriol. 178 (11), 3389-3395 (1996)
MEDLINE 96236067
REFERENCE 2 (bases 1 to 3196)
AUTHORS Strohl,W.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-1996) William R. Strohl, Ohio State University, 484 West 12th Avenue, Columbus, OH 43210 USA

```
FEATURES             Location/Qualifiers
     source          1. .3196
                     /organism="Streptomyces sp."
                     /strain="C5"
                     /db_xref="taxon:1931"
     CDS             <3. .590
                     /function="unknown ketoreductase"
                     /note="orf1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB08047.1"
                     /db_xref="GI:1354267"
                     /translation="YPRIDVMAGNAGGMFWSRTTTQDGFEATLQVNHLAGFLLARLLR
                     ERLAGGRLILTSSDAYTQGRIDPDDLNGDRHRYSAGQAYGTSKQANIMTATEAARRWP
                     DVLTVSYHPGEVRTRIGRGTVASTYFRFNPFLRSAAKGADTLVWLAAAPAEELTTGGY
                     YSDRRLSPVSGPTADAGLAAKLWEASAAAVGDTAR"
     gene            663. .1501
                     /gene="orfA"
     RBS             663. .666
                     /gene="orfA"
     CDS             674. .1501
                     /gene="orfA"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB08048.1"
                     /db_xref="GI:1354268"
                     /translation="MTRFAPGAPAWFDLGSPDVAASADFYTGLFGWTATVVSDPGAGG
                     YTTFSSDGKLVAAVARHQIDTPYHRPYGPGNDQHGMPAIWTVYFATNDADALTKRVET
                     AGGDVIMTPMDVLGLGRMAVFADPSGAAFAVWRKGVMEGAEVTGVPGSVGWVELVTDD
                     IGTARGFYRATLGLAPADTGRKGVTDPVWHIHDTPVAGTRELGTTGAVRPHWAVLFSV
                     HDCDATVRRAVELGGSVENEPVDTPRGRRADLLDPHGAGFSVVELREAYPAAADGAS"
     gene            1498. .2766
                     /gene="doxA"
     CDS             1498. .2766
                     /gene="doxA"
                     /function="hydroxylation of C-14 of daunomycin to produce doxorubicin"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="daunomycin C-14 hydroxylase"
                     /protein_id="AAB08049.1"
                     /db_xref="GI:1354269"
                     /translation="MSGEAPRVAVDPFSCPMMTMQRKPEVHDAFREAGPVVEVNAPAG
                     GPAWVITDDALAREVLADPRFVKDPDLAPTAWRGVDDGLDIPVPELRPFTLIAVDGED
                     HRRLRRIHAPAFNPRRLAERTDRIAAIADRLLTELADSSDRSGEPAELIGGFAYHFPL
                     LVICELLGVPVTDPAMAREAVGVLKALGLGGPQSAGGDGTDPAGDVPDTSALESLLLE
                     AVHAARRKDTRTMTRVLYERAQAEFGSVSDDQLVYMITGLIFAGHDTTGSFLGFLLAE
                     VLAGRLAADADGDAISRFVEEALRHHPPVPYTLWRFAATEVVIRGVRLPRGAPVLVDI
                     EGTNTDGRHHDAPHAFHPDRPSRRRLTFGDGPHYCIGEQLAQLESRTMIGVLRSRFPQ
                     ARLAVPYEELRWCRKGAQTARLTDLPVWLR"
     RBS             3038. .3041
                     /gene="dauI"
     gene            3038. .3196
                     /gene="dauI"
     CDS             3050. .>3196
                     /gene="dauI"
                     /function="putative transcriptional activator"
                     /codon_start=1
                     /transl_table=11
                     /product="DauI"
                     /protein_id="AAB08050.1"
                     /db_xref="GI:1354270"
                     /translation="MQINMLGPLVAHHNGTSVTPIARKPRQVFSLLALQAGTVVPVPA
                     LMEEL"
BASE COUNT 476 a 1184 c 1079 g 457 t
ORIGIN
```

Query Match 89.5%; Score 2567.8; DB 1; Length 3196;

Best Local Similarity 94.1%; Pred. No. 9.5e-205;
Matches 2697; Conservative 0; Mismatches 147; Indels 21; Gaps 2;

```
Qy       5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
Db    2845 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2786

Qy      65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
Db    2785 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2726

Qy     124 gggccccctccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
Db    2725 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2666

       184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
      2665 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2606

Qy     244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
Db    2605 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2546

Qy     304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
Db    2545 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2486

Qy     364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
Db    2485 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2426

Qy     424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
Db    2425 ACAACGTGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2366

Qy     484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
Db    2365 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2306

Qy     544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
Db    2305 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2246

Qy     604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
Db    2245 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2186

Qy     664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
Db    2185 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 2126

Qy     724 aggtgtccggcacgccccggcagggtccgtgccgtcaccccgccgctctgcgggccgc 783
Db    2125 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 2066

Qy     784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
Db    2065 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 2006

Qy     844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
Db    2005 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1946

Qy     904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
Db    1945 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1886

Qy     964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
Db    1885 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1826

Qy    1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
Db    1825 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1766

Qy    1084 gctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcgaggt 1143
Db    1765 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1706

Qy    1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
Db    1705 CGGGGTCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1646

Qy    1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
Db    1645 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 1586

Qy    1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
Db    1585 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1526

Qy    1324 cggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtacccctc 1383
Db    1525 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCGTCCGCCGCCGCGGGGTACGCCTC 1466

Qy    1384 ccgcagttcgaccaccgagaagccggcccgtgcgggtcgagcaggtccgcccgccgccc 1443
Db    1465 CCGCAGTTCGACCACCGAGAAGCCGGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCGCCC 1406

Qy    1444 cctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccggac 1503
Db    1405 CCTGGGGGTGTCGACGGGCTCGTTCTCGACGGAGCCGCCGAGTTCGACGGCCCGCCGGAC 1346

Qy    1504 cgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgccggt 1563
Db    1345 CGTCGCGTCGCAGTCGTGCACGGAGAACAGCACGGCCCAGTGGGGCCGTACCGCGCCGGT 1286

Qy    1564 gacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtcggt 1623
Db    1285 CGTGCCCAGTTCCCGGGTGCCGGCGACCGGTGTGTCATGGATGTGCCAAACCGGGTCGGT 1226

Qy    1624 gacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtcccg 1683
Db    1225 GACGCCCTTGCGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCACGGTAGAAGCCACG 1166

Qy    1684 ggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcccgt 1743
Db    1165 GGCGGTCCCGATGTCGTCGGTCACCAGTTCCACCCAGCCGACCGAGCCGGGCACGCCCGT 1106

Qy    1744 cacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggggtc 1803
Db    1105 CACCTCCGCGCCCTCCATGACGCCCTTGCGCCACACCGCGAACGCGGCCCCCGATGGGTC 1046

Qy    1804 ggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctcgcc 1863
Db    1045 GGCGAAGACCGCCATCCGGCCGAGACCGAGGACGTCCATCGGGGTCATGATGACGTCGCC 986

Qy    1864 gcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacacggt 1923
Db     985 ACCCGCCGTCTCGACCCGTTTGGTCAGTGCGTCGGCGTCGTTGGTGGCGAAGTACACGGT 926

Qy    1924 ccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggggct 1983
Db     925 CCAGATGGCCGGCATGCCGTGCTGGTCGTTGCCGGGCCCGTACGGACGGTGGTAGGGCGT 866

Qy    1984 gtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgtgta 2043
Db     865 GTCGATCTGATGGCGGGCGACCGCGGCGACCAGCTTCCCGTCGGAGCTGAACGTCGTGTA 806

Qy    2044 tccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgtagaa 2103
Db     805 TCCCCCGGCACCCGGGTCGCTGACCACGGTCGCGGTCCAGCCGAAGAGGCCGGTGTAGAA 746

Qy    2104 gtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccgggcgc 2163
Db     745 GTCGGCCGAGGCGGCGACATCGGGCGACCCGAGGTCGAACCATGCGGGGGCGCCGGGCGC 686
```

```
Qy   2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632

Qy   2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
               ||||||||||||||| |||||||||||| || ||||||| |||||
Db    633 ------------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGGCCCGCCGCGGTCAGC 586

Qy   2284 gcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccggcgt 2343
          ||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||
Db    585 GCGCGGTGTCGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526

Qy   2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCA 466

Qy   2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
          |||||||||||||||||| |||||||||||||| |||||||||||| |||||||| |
Db    465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406

Qy   2464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
          |||||||||||||||||||||||||| ||||||||| |||||||| ||||||||||||
Db    405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346

Qy   2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
          ||||||| |||||||| |||||||||| |||||||||||||||||| ||||||||||
Db    345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286

Qy   2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
            ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226

Qy   2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166

Qy   2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||| |
Db    165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCCG 106

Qy   2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
          |||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46

Qy   2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868
          |||||||||||||||||| |||||||||||||||||||||||||
Db     45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1
```

RESULT 3
[redacted]78146/c
[redacted]S AR078146 3196 bp DNA linear PAT 31-AUG-2000
[redacted]NITION Sequence 4 from patent US 5962293.
[redacted]SSION AR078146
VERSION AR078146.1 GI:10004892
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 3196)
AUTHORS Strohl,W.R., Dickens,M.L. and Desanti,C.L.
TITLE Methods of producing daunorubicin
JOURNAL Patent: US 5962293-A 4 05-OCT-1999;
FEATURES Location/Qualifiers
source 1. .3196
/organism="unknown"
BASE COUNT 475 a 1185 c 1076 g 460 t
ORIGIN

Query Match 88.6%; Score 2542.2; DB 6; Length 3196;
Best Local Similarity 93.6%; Pred. No. 1.3e-202;

Matches 2681; Conservative 0; Mismatches 163; Indels 21; Gaps 2;

```
Qy      5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
          |||||||||| |||||||||||||||||||||||| || ||||||||||||||||||||
Db   2845 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2786

Qy     65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
          || |||||||| |||| |||| |||||||||||||||| || |||||||||| ||||
Db   2785 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2726

Qy    124 gggccccctccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
          |||||||||||| |||||||||||||| |||||||||||||||||||||||| | |||||
Db   2725 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2666

Qy    184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   2665 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2606

Qy    244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
          |||||||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||
Db   2605 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2546

Qy    304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2545 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2486

Qy    364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
          |||||||||||||||||||||||| |||| ||||  ||||||||||||||||||||||||
Db   2485 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2426

Qy    424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   2425 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2366

Qy    484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
          ||||| |||||| |||||||||| |||||||||||||||||| ||||||||||||||||
Db   2365 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2306

Qy    544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2305 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2246

Qy    604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
          |||||||||||||||||||||||| ||||| ||||| || || ||||||||||||||||
Db   2245 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2186

Qy    664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
          || ||||||| |||||||||| | ||||||||||||||||| |||||||||| ||||
Db   2185 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 2126

Qy    724 aggtgtccggcacgcccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
          | |||||||||||| ||||||||||||||||||||||||||||| ||| ||||||||||||
Db   2125 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 2066

Qy    784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
          |||||||||||||||||| |||| |||||||||||||||||| ||||||||||||||||
Db   2065 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 2006

Qy    844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
          |||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   2005 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1946

Qy    904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
          |||||||||||| |||||||||| |||||| |||||||||||||||||||||||||| |||
Db   1945 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1886

Qy    964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
          |||| ||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   1885 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1826

Qy   1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
          |||||||||||| ||||||||| |||| ||||||||||||||||||||||||||||||||
Db   1825 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1766
```

```
Qy   1084 gctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcgaggt 1143
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |
Db   1765 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCCGCCAGGCGGTGGGCGCGAGAT 1706

Qy   1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
          ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1705 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1646

Qy   1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||
Db   1645 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 1586

Qy   1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
          | |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
     1585 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1526

     1324 cggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtacccctc 1383
          |||||||||| |||||||||||||||| ||||||||| |||||||||||||||||| ||||
Db   1525 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCGTCCGCCGCCGCGGGGTACGCCTC 1466

Qy   1384 ccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccgccc 1443
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1465 CCGCAGTTCGACCACCGAGAAGCCGGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCGCCC 1406

Qy   1444 cctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccggac 1503
          |||||| |||||| |||||||||||||||||||||||||||||||| ||||| ||||||||
Db   1405 CCTGGGGGTGTCGACGGGCTCGTTCTCGACGGAGCCGCCGAGTTCGACGGCCCGCCGGAC 1346

Qy   1504 cgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgccggt 1563
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   1345 CGTCGCGTCGCAGTCGTGCACGGAGAACAGCACGGCCCAGTGGGGCCGTACCGCGCCGGT 1286

Qy   1564 gacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtcggt 1623
          ||||||| ||| |||||||||||||||||||||||||| |||||||||| ||||||||||
Db   1285 CGTGCCCAGTTCCCGGGTGCCGGCGACCGGTGTGTCATGGATGTGCCAAACCGGGTCGGT 1226

Qy   1624 gacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtcccg 1683
          |||||||||  ||||||||||||||||||||||||||||||||||  |||||||| | ||
Db   1225 GACGCCCTTGCGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCACGGTAGAAGCCACG 1166

Qy   1684 ggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcccgt 1743
          ||||| |||||||  |||||||||||||| || |||||||||||||||||||||||||||||
Db   1165 GGCGGTCCCGATGTCGTCGGTCACCAGTTCCACCCAGCCGACCGAGCCGGGCACGCCCGT 1106

     1744 cacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggggtc 1803
          |||||||||||||||||||||  |||||||||||||||| |||||||||||||||| |  |||||
Db   1105 CACCTCCGCGCCCTCCATGACGCCCTTGCGCCACACCGCGAACGCGGCCCCCGATGGGTC 1046

Qy   1804 ggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctcgcc 1863
          |||||||||||||||||||||| ||||||||||||||||   ||||||||| |||||
Db   1045 GGCGAAGACCGCCATCCGGCCGAGACCGAGGACGTCCATCGGGTGCATGATGACGTCGCC 986

Qy   1864 gcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacacggt 1923
          |||||||  |||||||| |||||||||||||||||||||| ||||||||||||||||||
Db    985 ACCCGCCGCTTCGACCCGTTTGGTCAGTGCGTCGGCGTCGTTGGTGGCGAAGTACACGGT 926

Qy   1924 ccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtagggggt 1983
          |||||||||||||||||||||||  || ||||||||||||| ||||||||||| ||
Db    925 CCAGATGGCCGGCATGCCGTGCTGGTGCTTGCCGGGCCCGTACGGACGGTGGTAGGGCGT 866

Qy   1984 gtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgtgta 2043
          |||||||| |||||||||||||||||  |||||||||||||||||| || |||||
Db    865 GTCGATCTGATGGCGGGCGACCGCGGCGACAGGCTTCCCGTCGGAGCTGAAAGTAGTGTA 806

Qy   2044 tcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgtagaa 2103
          |||||||| |||||||||||||||||||| |||||||| |||| ||||||||||||||
Db    805 TCCCCCGGCACCCGGGTCGCTGACCACGGTCGCGGTCCACGCGAAGAGGCCGGTGTAGAA 746

Qy   2104 gtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccgggcgc 2163
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    745 GTCGGCCGAGGCGGCGACATCGGGCGACCCGAGGTCGAACCATGCGGGGGCGCCGGGCGC 686
```

```
Qy   2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632

Qy   2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
                   |||||||||||||||| |||||||||||| || ||||||| |||||
Db    633 ------------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGGCCCGCCGCGGTCAGC 586

Qy.  2284 gcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccggcgt 2343
          ||||||||| |||||||||||||||| |||||||||||||| ||||||||||||||||
Db    585 GCGCGGTGTGGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526

Qy   2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCA 466

Qy   2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
          |||||||||||||||| |||||||||||||| |||||||||||||| |||||||| |
Db    465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406

Qy   2464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
          |||||||||||||||||||||||||||||| ||||||||| |||||||| ||||||||||||
Db    405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346

Qy   2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
          ||||||| |||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286

Qy   2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
            |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226

Qy   2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166

Qy   2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
          |||||||||  |||||||||||||||||||||||||||||||||||||||||||||||| |
Db    165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCCG 106

Qy   2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
          |||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46

Qy   2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868
          |||||||||||||||||| ||||||||||||||||||||||||||
Db     45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1
```

```
RESULT   4
AR083230/c
LOCUS       AR083230               3196 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 4 from patent US 5976830.
ACCESSION   AR083230
VERSION     AR083230.1  GI:10010020
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3196)
  AUTHORS   Strohl,W.R., Dickens,M.L. and Desanti,C.L.
  TITLE     Methods of producing doxorubicin
  JOURNAL   Patent: US 5976830-A 4 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .3196
                     /organism="unknown"
BASE COUNT      475 a   1185 c   1076 g    460 t
ORIGIN
```

```
  Query Match             88.6%;  Score 2542.2;  DB 6;  Length 3196;
  Best Local Similarity   93.6%;  Pred. No. 1.3e-202;
  Matches 2681;  Conservative    0;  Mismatches  163;  Indels   21;  Gaps    2;
```

```
Qy       5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
Db    2845 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2786

Qy      65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
Db    2785 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2726

Qy     124 gggcccccttccggcaccaccgcaactcgtcgtacggcacggccagtcggcctcgggga 183
Db    2725 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGA 2666

Qy     184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
Db    2665 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2606

Qy     244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
Db    2605 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2546

Qy     304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
Db    2545 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2486

Qy     364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
Db    2485 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2426

Qy     424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
Db    2425 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2366

Qy     484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcagcaggaagccca 543
Db    2365 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2306

Qy     544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
Db    2305 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2246

Qy     604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
Db    2245 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2186

Qy     664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
Db    2185 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 2126

Qy     724 aggtgtccggcacgcccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
Db    2125 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 2066

Qy     784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
Db    2065 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 2006

Qy     844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
Db    2005 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1946

Qy     904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
Db    1945 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1886

Qy     964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
Db    1885 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1826

Qy    1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
Db    1825 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1766
```

```
Qy    1084 gctccggaacggggatgtcgagaccgtcgtccaccccccgccaggcggcgggggcgaggt 1143
Db    1765 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1706

Qy    1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
Db    1705 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1646

Qy    1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccggcccgcctcccgga 1263
Db    1645 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGCCCGCCTCTCGGA 1586

Qy    1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
Db    1585 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1526

Qy    1324 cggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtacccctc 1383
Db    1525 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCGTCCGCCGCCGCGGGGTACGCCTC 1466

Qy    1384 ccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccgccc 1443
Db    1465 CCGCAGTTCGACCACCGAGAAGCCGGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCGCCC 1406

Qy    1444 cctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccggac 1503
Db    1405 CCTGGGGGTGTCGACGGGCTCGTTCTCGACGGAGCCGCCGAGTTCGACGGCCCGCCGGAC 1346

Qy    1504 cgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgccggt 1563
Db    1345 CGTCGCGTCGCAGTCGTGCACGGAGAACAGCACGGCCCAGTGGGGCCGTACCGCGCCGGT 1286

Qy    1564 gacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtcggt 1623
Db    1285 CGTGCCCAGTTCCCGGGTGCCGGCGACCGGTGTGTCATGGATGTGCCAAACCGGGTCGGT 1226

Qy    1624 gacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtcccg 1683
Db    1225 GACGCCCTTGCGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCACGGTAGAAGCCACG 1166

Qy    1684 ggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcccgt 1743
Db    1165 GGCGGTCCCGATGTCGTCGGTCACCAGTTCCACCCAGCCGACCGAGCCGGGCACGCCCGT 1106

Qy    1744 cacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggggtc 1803
Db    1105 CACCTCCGCGCCCTCCATGACGCCCTTGCGCCACACCGCGAACGCGGCCCCCGATGGGTC 1046

Qy    1804 ggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctcgcc 1863
Db    1045 GGCGAAGACCGCCATCCGGCCGAGACCGAGGACGTCCATCGGGTGCATGATGACGTCGCC 986

Qy    1864 gcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacacggt 1923
Db     985 ACCCGCCGCTTCGACCCGTTTGGTCAGTGCGTCGGCGTCGTTGGTGGCGAAGTACACGGT 926

Qy    1924 ccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtagggggt 1983
Db     925 CCAGATGGCCGGCATGCCGTGCTGGTGCTTGCCGGGCCCGTACGGACGGTGGTAGGGCGT 866

Qy    1984 gtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgtgta 2043
Db     865 GTCGATCTGATGGCGGGCGACCGCGGCGACAGGCTTCCCGTCGGAGCTGAAAGTAGTGTA 806

Qy    2044 tcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgtagaa 2103
Db     805 TCCCCCGGCACCCGGGTCGCTGACCACGGTCGCGGTCCACGCGAAGAGGCCGGTGTAGAA 746

Qy    2104 gtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccgggcgc 2163
Db     745 GTCGGCCGAGGCGGCGACATCGGGCGACCCGAGGTCGAACCATGCGGGGGCGCCGGGCGC 686

Qy    2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
```

```
Db    685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632

Qy   2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
               |||||||||||||||| |||||||||||| || ||||||| |||||
Db    633 -----------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGGCCCGCCGCGGTCAGC 586

Qy   2284 gcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccggcgt 2343
          ||||||||| |||||||||||||||||| |||||||||||||| ||||||||||||||||
Db    585 GCGCGGTGTGGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526

Qy   2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCA 466

     2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
          |||||||||||||||| |||||||||||||| |||||||||||||| |||||||| |
      465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406

Qy   2464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
          |||||||||||||||||||||||||| ||||||||| |||||||| ||||||||||||||
Db    405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346

Qy   2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
          ||||||| |||||||| |||||||||| ||||||||||||||||||| ||||||||||||
Db    345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286

Qy   2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226

Qy   2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166

Qy   2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||| |
Db    165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCCG 106

Qy   2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
          |||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46

Qy   2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868
          |||||||||||||||| |||||||||||||||||||||||||||
Db     45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1
```

LT 5
AF403708/c

```
LOCUS       AF403708               1932 bp    DNA     linear   BCT 29-AUG-2001
DEFINITION  Streptomyces peucetius subsp. caesius daunomycin C-14 hydroxylase
            (doxA) gene, complete cds.
ACCESSION   AF403708
VERSION     AF403708.1  GI:15341421
KEYWORDS    .
SOURCE      Streptomyces peucetius subsp. caesius.
  ORGANISM  Streptomyces peucetius subsp. caesius
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 1932)
  AUTHORS   Hong,Y.-S., Kim,H.S., Lee,J.-H., Kim,K.-W. and Lee,J.J.
  TITLE     Molecular Cloning and Characterization of the doxA Cytochrome P-450
            Hydroxylase Gene in Streptomyces peucetius subsp. caesius ATCC
            27952
  JOURNAL   J. Microbiol. Biotechnol. (2001) In press
REFERENCE   2  (bases 1 to 1932)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2001) Anticancer Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, YuSong, P. O.
            Box 115, Taejon 305-600, Korea
FEATURES             Location/Qualifiers
     source          1. .1932
                     /organism="Streptomyces peucetius subsp. caesius"
                     /strain="ATCC 27952"
                     /sub_species="caesius"
                     /db_xref="ATCC:27952"
                     /db_xref="taxon:55158"
     gene            591. .1853
                     /gene="doxA"
     CDS             591. .1853
                     /gene="doxA"
                     /note="cytochrome P-450 hydroxylase"
                     /codon_start=1
                     /transl_table=11
                     /product="daunomycin c-14 hydroxylase"
                     /protein_id="AAK95626.1"
                     /db_xref="GI:15341422"
                     /translation="MSGEAPRVAVDPFACPMMTMQRKPEVHDAFREAGPVVEVNAPAG
                     GPAWFITDDALSRYVLADPRLVKDPDLAPAAWRGVVDGLDIPVPELRPFTLIAVDGEA
                     HRRLHRIHAPAFNPRRLAERTDRIAAIAGRLLTELADASGRSGEPAELIGGFAYHFPL
                     LVICELLGVPVTVPMAREAVSVLKALASAAQSGGGDGTDPAGGVPDTSALESLLLEAV
                     HSARRNDTPTMTRVLYEHTQAEFGSVSDNQLVYMITGIIFAGHERTGSFLGFLLAEVL
                     AGRLAADADEDAVSRFVEEAVRYHPPVPYTLWRFAATEVTIGGVRLPPGAPVLVDIEG
                     TNTDGRHHDAPHAFHPDRPSWRRLTFGDGPHYCIGEQLAQLESRTMIGVLRSRFPEAR
                     LAVPYDELRWCRNGAQTARLTELPVWLR"
BASE COUNT      250 a    714 c    690 g    278 t
ORIGIN
```

```
  Query Match             60.4%;  Score 1732.4;  DB 1;  Length 1932;
  Best Local Similarity   96.3%;  Pred. No. 1.7e-135;
  Matches 1872;  Conservative    0;  Mismatches   56;  Indels   16;  Gaps    9;

Qy      1 ggatccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||  |   ||  |
Db   1932 GGATCCGCACCGGGTACACGGCACGGGACCGCCCACCGCGCGGTGCGGTGGTCTCG---C 1876

Qy     61 cgtgccggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccg 120
           | ||    | ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1875 TGCGCTCCGCCCGGCCGGCGGATCAGCGCAGCCAGACGGGCAGTTCGGTGAGCCGCGCCG 1816

Qy    121 tctgggccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgg 180
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1815 TCTGGGCCCCGTTCCGGCACCACCGCAACTCGTCGTACGGCACGGCCAGTCGGGCCTCGG 1756

Qy    181 ggaacctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1755 GGAACCTGCTGCGCAGTACGCCGATCATCGTGCGCGACTCCAGCTGGGCGAGCTGCTCCC 1696

Qy    241 cgatgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1695 CGATGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCACGAGGGACGGTCCGGGT 1636

Qy    301 ggaaggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccacca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1635 GGAAGGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCA 1576

Qy    361 gcaccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacc 420
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1575 GCACCGGCGCTCCGGGGGGCAGCCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACC 1516

Qy    421 tccacaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggaga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1515 TCCACAACGTGTAGGGCACCGGCGGGTGGTAGCGCACCGCCTCCTCCACGAACCGGGAGA 1456

Qy    481 cggcgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1455 CGGCGTCCTCGTCGGCATCCGCCGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGC 1396

Qy    541 ccaggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacga 600
          |||||||||||||||  | |||||||||||||||||| ||||||||||||||||||||
Db   1395 CCAGGAAGGAGCCGGTCCTCTCGTGGCCGGCGAAGATGATCCCGGTGATCATGTAGACGA 1336

Qy    601 gctggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtca 660
```

```
Db   1335 GCTGGTTGTCGGAGACCGAGCCGAACTCGGCCTGCGTGTGCTCGTACAGCACGCGGGTCA 1276

Qy    661 tggtcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccaggg 720

Db   1275 TGGTCGGGGTGTCGTTCCGCCGGGCTGAGTGCACGGCTTCGAGGAGCAGGCTCTCCAGGG 1216

Qy    721 ccgaggtgtccggcacgcccccggcagggtccgtgccgtcacccccgccgctctgcgggc 780

Db   1215 CCGAGGTGTCCGGCACGCCCCCGGCAGGGTCCGTGCCGTCACCCCCGCCGCTCT--GGGC 1158

Qy    781 cgccgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtga 840

Db   1157 CGCCGAGG-CGAGTGCCTTGAGAACGCTGACGGCCTCGCGGGCCAT---CGGAACGGTGA 1102

Qy    841 ccggcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccga 900

Db   1101 CCGGCACACCGAGCAGCTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGA 1042

Qy    901 tcagctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggc 960

Db   1041 TCAGCTCGGCCGGTTCGCCCGACCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGC 982

Qy    961 cggcgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtg 1020

Db    981 CGGCGATCGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTG 922

Qy   1021 cgtggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggac 1080

Db    921 CGTGGATGCGGTGCAGGCGCCGGTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGAC 862

Qy   1081 gcagctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcga 1140

Db    861 GCAGCTCCGGGACGGGGATGTCGAGACCGTCGACCACCCCCCGCCAGGCGGCGGGGGCGA 802

Qy   1141 ggtcggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcgg 1200

Db    801 GGTCGGGGTCCTTCACCAACCGGGGATCGGCCAGCACATAGCGGGAGAGGGCGTCATCGG 742

Qy   1201 tgatgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctccc 1260

Db    741 TGATGAACCAGGCGGGTCCGCCAGCAGGGGCGTTCACCTCAACGACCGGGCCCGCCTCCC 682

Qy   1261 ggaaggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggt 1320

Db    681 GGAAGGCGTCGTGCACCTCGGGCTTGCGCTGCATGGTCATCATGGGACACGCGAACGGGT 622

Qy   1321 cgacggccacccggggcgcctcgccgctcacgaggcacc-gcccgccgccgcggggtacc 1379

Db    621 CGACGGCCACCCGGGGCGCCTCGCCGCTCACGAGGCACCTGCCCGCCGCCGCGGGGTACC 562

     1380 cctcccgcagttcgaccaccgagaagccggccccgtgcgggtc-gagcaggtccgcccgc 1438

      561 CCTCCCGCAGTTCGACCACCGAGAAGCCGCCCCCGTACGGGTCAGAGCAGGTCCGCCCAC 502

Qy   1439 cgccccctgggcgt-gtcggcgggctcgttctcgacggagccgccgagttcaacggcgcg 1497

Db    501 CGCCCCCTGGGCGTAGTCGGCGGGCTCGTTCTCGACGGAGCCGCCGAGTTCAACGGCGCG 442

Qy   1498 ccggaccgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgc 1557

Db    441 CCGGACCGTCGCGTCGCAGTCGTGCACGGCGAACAGCACGGCCCAGTGCGGCCGTACCGC 382

Qy   1558 gccggtgacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgg 1617

Db    381 GCCGGTGACGCCCAGCTCCTGGGTGCCGGCGACCGGTGTGTGACCGATGTGCCAGACCGG 322

Qy   1618 gtcggtgacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaa 1677

Db    321 GTCAATGACGCCCTTCAGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCCGGGTAGAA 262

Qy   1678 gtcccgggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcac 1737

Db    261 GTCCCGGGCGGCCCCGATGCCGTCGGTCACCAGCTCGACCCAGCCGACCGAGCCGGGCAC 202

Qy   1738 gcccgtcacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggc 1797

Db    201 GCCCGTCACCTCCGCGCCCTCCATGACTCCCTTGCGCCAGAGACGCGAACGGGCC---GG 145

Qy   1798 ggggtcggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgac 1857

Db    144 CGGGGTGGCGAAGACCGCCATCCGGCCGAGGCCGAGGACGTCCATCGGAGTCATGATGAC 85

Qy   1858 ctcgccg-cccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagt 1916

Db     84 CTCGCCGCCCCGCCGTCTCGACCCGCTTGGTCAGTGCGTCGGCGTCGTCGGTGGCGAAGT 25

Qy   1917 acacggtccagatggccggcatgc 1940

Db     24 ACACGGTCTAGATGGCCGGCATGC 1
```

RESULT 6
I84584/c
LOCUS I84584 1269 bp DNA linear PAT 04-APR-1998
DEFINITION Sequence 1 from patent US 5695966.
ACCESSION I84584
VERSION I84584.1 GI:3022104
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1269)
AUTHORS Inventi,A.Solari., Breme,U., Colombo,A.Luisa., Hutchinson,C.Richard., Otten,S. and Scotti,C.
TITLE DNA encoding daunorubicin 14-hyroxylase and method for preparing doxorubicin
JOURNAL Patent: US 5695966-A 1 09-DEC-1997;
FEATURES Location/Qualifiers
source 1. .1269
/organism="unknown"
BASE COUNT 164 a 484 c 442 g 179 t
ORIGIN

Query Match 44.2%; Score 1269; DB 6; Length 1269;
Best Local Similarity 100.0%; Pred. No. 4.5e-97;
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     83 tcagcgcagccagacgggcagttcggtgagccgcgccgtctgggccccctccggcacca 142

Db   1269 TCAGCGCAGCCAGACGGGCAGTTCGGTGAGCCGCGCCGTCTGGGCCCCCTTCCGGCACCA 1210

Qy    143 ccgcaactcgtcgtacggcacggccagtcgggcctcggggaacctgctgcgcagtacgcc 202

Db   1209 CCGCAACTCGTCGTACGGCACGGCCAGTCGGGCCTCGGGGAACCTGCTGCGCAGTACGCC 1150

Qy    203 gatcatcgtgcgcgactccagctgggcgagctgctccccgatgcagtagtgcggcccgtc 262

Db   1149 GATCATCGTGCGCGACTCCAGCTGGGCGAGCTGCTCCCCGATGCAGTAGTGCGGCCCGTC 1090

Qy    263 gccgaaggtgagccgccgccacgagggacggtccgggtggaaggcgtgcggggcgtcgtg 322

Db   1089 GCCGAAGGTGAGCCGCCGCCACGAGGGACGGTCCGGGTGGAAGGCGTGCGGGGCGTCGTG 1030

Qy    323 atggcggccgtcggtgttggtgccctcgatgtccaccagcaccggcgctccgcggggcag 382

Db   1029 ATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCCGCGGGGCAG 970

Qy    383 ccggacgccgccgatggtcacctccgtggcagcgaacctccacaacgtgtagggcaccgg 442

Db    969 CCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACCTCCACAACGTGTAGGGCACCGG 910

Qy    443 cgggtggtagcgcagcgcctcctccacgaaccgggagacggcgtcctcgtcggcatccgc 502

Db    909 CGGGTGGTAGCGCAGCGCCTCCTCCACGAACCGGGAGACGGCGTCCTCGTCGGCATCCGC 850
```

```
Qy    503 cgcgaggcggcccgccaggacctccgcgagcaggaagcccaggaaggagccggtggtgtc 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 CGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGGAGCCGGTGGTGTC 790

Qy    563 gtggccggcgaagatgagcccggtgatcatgtagacgagctggtcgtcggagaccgagcc 622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 GTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACGAGCTGGTCGTCGGAGACCGAGCC 730

Qy    623 gaactcggcctgcgcgcgctcgtacagcacgcgggtcatggtcggggtgtcgttccgccg 682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCATGGTCGGGGTGTCGTTCCGCCG 670

Qy    683 ggctgagtgcacggcttcgaggagcaggctctccagggccgaggtgtccggcacgccccc 742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      669 GGCTGAGTGCACGGCTTCGAGGAGCAGGCTCTCCAGGGCCGAGGTGTCCGGCACGCCCCC 610

      743 ggcagggtccgtgccgtcacccccgccgctctgcgggccgccgaggccgagtgccttgag 802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 GGCAGGGTCCGTGCCGTCACCCCCGCCGCTCTGCGGGCCGCCGAGGCCGAGTGCCTTGAG 550

Qy    803 aacgctgacggcctcgcgggccatcgccggatcggtgaccggcacaccgagcagctcgca 862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 AACGCTGACGGCCTCGCGGGCCATCGCCGGATCGGTGACCGGCACACCGAGCAGCTCGCA 490

Qy    863 gatgaccaacagcgggaagtggtacgcgaagccgccgatcagctcggccggtttgcccga 922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 GATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCAGCTCGGCCGGTTTGCCCGA 430

Qy    923 ccggccggaggcgtcggcgagttcggtgagcagccggccggcgatcgcggcgatgcgatc 982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 CCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCCGGCGATCGCGGCGATGCGATC 370

Qy    983 cgtccgctcggccagccggcgcgggttgaacgcaggtgcgtggatgcggcgcaggcgccg 1042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTGCGTGGATGCGGCGCAGGCGCCG 310

Qy   1043 gtgggcctcgccgtccacggcgatgagcgtgaacggacgcagctccggaacggggatgtc 1102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 GTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGACGCAGCTCCGGAACGGGGATGTC 250

Qy   1103 gagaccgtcgtccacccccccgccaggcggcgggggcgaggtcggggtccttcacgaaccg 1162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 GAGACCGTCGTCCACCCCCCCGCCAGGCGGCGGGGGCGAGGTCGGGGTCCTTCACGAACCG 190

Qy   1163 gggatcggccagcacctcgcgggcgagggcgtcatcggtgatgacccaggcgggtccgcc 1222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 GGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGATGACCCAGGCGGGTCCGCC 130

Qy   1223 cgcgggggcgttcacctcgacgaccgggcccgcctcccggaaggcgtcgtgcacctcggg 1282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 CGCGGGGGCGTTCACCTCGACGACCGGGCCCGCCTCCCGGAAGGCGTCGTGCACCTCGGG 70

Qy   1283 cttgcgctgcatggtcatcatgggacacgcgaacgggtcgacggccacccggggcgcctc 1342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 CTTGCGCTGCATGGTCATCATGGGACACGCGAACGGGTCGACGGCCACCCGGGGCGCCTC 10

Qy   1343 gccgctcac 1351
          |||||||||
Db       9 GCCGCTCAC 1
```

```
RESULT   7
AR078148/c
LOCUS       AR078148               2081 bp    DNA     linear   PAT 31-AUG-2000
DEFINITION  Sequence 7 from patent US 5962293.
ACCESSION   AR078148
VERSION     AR078148.1  GI:10004894
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2081)
  AUTHORS   Strohl,W.R., Dickens,M.L. and Desanti,C.L.
  TITLE     Methods of producing doxorubicin
  JOURNAL   Patent: US 5962293-A / 05-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .2081
                     /organism="unknown"
BASE COUNT      380 a    688 c    657 g    356 t
ORIGIN
```

Query Match 42.6%; Score 1221.6; DB 6; Length 2081;
Best Local Similarity 94.1%; Pred. No. 3.1e-93;
Matches 1280; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

```
Qy      5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
          ||||||||||    |||||||||||||||||||||||| || ||||||||||||||||||
Db   1730 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 1671

Qy     65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
          || |||||||| ||||  |||| |||||||||||||||| || ||||||||||| ||||
Db   1670 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 1611

Qy    124 gggccccctttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
          |||||||||||| |||||||||||||| |||||||||||||||||||||||| | |||||
Db   1610 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 1551

Qy    184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1550 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 1491

Qy    244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
          |||||||||||||||||||||||||||||||||||| ||| || ||||||||||||||||
Db   1490 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 1431

Qy    304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   1430 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 1371

Qy    364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
          |||||||||||||||||||||||| |||| |||| ||||||||||||||||||||||||
Db   1370 CCGGCGCTCCGCGGGGCAGCCGGACACCGCCGGATGACCACCTCCGTGGCAGCGAACCTCC 1311

Qy    424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   1310 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 1251

Qy    484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
          ||||| ||||||| |||||||||| ||||||||||||||||||| ||||||||||||||
Db   1250 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 1191

Qy    544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
          |||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1190 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 1131

Qy    604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
          |||||||||||||||||||||||| ||||| ||||| || || ||||||||||||||||
Db   1130 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 1071

Qy    664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
          || ||||||| |||||||||| | ||||||||||||||||| |||||||||| ||||||
Db   1070 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 1011

Qy    724 aggtgtccggcacgcccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
          | |||||||||||| |||||||||||||||||||||||||| ||| ||||||||||||||
Db   1010 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 951

Qy    784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
          |||||||||||||||||| |||| ||||||||||||||||| ||||||||||||||||||
Db    950 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 891

Qy    844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
          |||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    890 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 831
```

```
Qy    904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
          ||||||||||||| |||||||||| |||||| |||||||||||||||||||||||||| |||
Db    830 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 771

Qy    964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    770 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 711

Qy   1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
          |||||||||||| ||||||||| |||| ||||||||||||||||||||||||||||||||
Db    710 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 651

Qy   1084 gctccggaacggggatgtcgagaccgtcgtccacccccgccaggcggcgggggcgaggt 1143
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |
Db    650 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 591

Qy   1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
          ||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 531

Qy   1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||
Db    530 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 471

Qy   1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
          | |||||||||||||||| |||||||||||||||||||||||||| ||||||||||| |
Db    470 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 411

Qy   1324 cggccacccggggcgcctcgccgctcacgaggcaccgccc 1363
          |||||||||| |||||||||||||||| |||||||| | |
Db    410 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCCCTC 371
```

Search completed: June 11, 2002, 20:01:13
Job time: 7335 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 17:59:13 ; Search time 308.46 Seconds
(without alignments)
15974.663 Million cell updates/sec

Title: US-09-673-254-1
Perfect score: 2870
Sequence: 1 ggatccgcaccgggtacacg.....catcgatgcgcgggtaccgg 2870

[redacted]g table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2870	100.0	2870	21	AAZ31413	Complementary DNA
c 2	2558.2	89.1	3196	19	AAV01453	Streptomyces dauno
c 3	2542.2	88.6	3196	19	AAV01447	Daunomycin C-14 hy
c 4	1269	44.2	1269	17	AAT33617	DxrA gene encoding
c 5	1224.8	42.7	3013	19	AAV01451	Plasmid pANT195 in
c 6	1221.6	42.6	2081	19	AAV01452	Modified doxA gene
c 7	1044.4	36.4	1569	19	AAV34001	DNA fragment conta
8	182	6.3	114955	20	AAX53491	Human adenosine A1
c 9	181.2	6.3	125401	22	AAD17186	Streptomyces nours
c 10	164.4	5.7	114955	20	AAX53491	Human adenosine A1
c 11	154.8	5.4	2919	21	AAZ24486	Streptomyces tende
c 12	152.6	5.3	12249	21	AAC55840	Complete Mitomycin
c 13	152.6	5.3	18331	21	AAC55857	Complete nucleotid
c 14	152.2	5.3	2918	21	AAZ43922	S. tendae nikkomyc
c 15	142	4.9	65140	22	AAD17184	Streptomyces nours
c 16	139.8	4.9	1257	22	AAF81357	Quorum sensing con
c 17	138.6	4.8	1224	21	AAC55788	Hydroxylase orf4 i
c 18	137.4	4.8	4257	19	AAV68520	The nucleotide seq
c 19	137.4	4.8	4257	19	AAV10362	Infected cell prot
c 20	137.4	4.8	5970	21	AAA75635	Nucleotide sequenc
c 21	137.4	4.8	5970	21	AAZ56003	Contig 002 from co
c 22	133.8	4.7	1251	21	AAZ87301	S. venezuelae macr
c 23	133.2	4.6	4496	17	AAT58553	Streptomyces prist
24	132.6	4.6	12001	16	AAQ76213	HSV L/ST region.
c 25	130.6	4.6	44377	18	AAT78508	Platenolide syntha
c 26	130.6	4.6	44377	18	AAT80414	Platenolide syntha
c 27	128.6	4.5	1700	14	AAQ45569	Sequence soyC and
c 28	127.4	4.4	2795	22	AAI71296	Streptomyces grise
c 29	127.4	4.4	2795	22	AAF81954	Streptomyces grise
c 30	127.2	4.4	27541	22	AAD17185	Streptomyces nours
c 31	123.8	4.3	3957	22	AAA09686	HSV-2 immediate ea
32	123.8	4.3	154746	24	AAD25519	Human herpesvirus
c 33	123.8	4.3	154746	24	AAD25519	Human herpesvirus
c 34	122.2	4.3	109519	22	AAS08693	Micromonospora DNA
c 35	122	4.3	58857	21	AAA58471	Nucleotide sequenc
c 36	121.6	4.2	1194	17	AAT58555	Streptomyces prist
37	120.8	4.2	14806	22	AAD03809	Streptomyces galil
c 38	119	4.1	940	19	AAV63943	Mycobacterium tube
c 39	119	4.1	940	20	AAX81092	Nucleotide sequenc
c 40	117.6	4.1	24379	18	AAT93095	Streptomyces freno
c 41	117.6	4.1	24379	19	AAV25925	Streptomyces roseo
c 42	117	4.1	50937	21	AAA09469	Streptococcus olea
43	115.6	4.0	8438	15	AAQ73500	DNA encoding Pseud
c 44	114.6	4.0	1998	12	AAQ11127	Sequence encoding
c 45	113.8	4.0	1879	12	AAQ11126	Sequence encoding

ALIGNMENTS

RESULT 1
AAZ31413
ID AAZ31413 standard; DNA; 2870 BP.
XX
AC AAZ31413;
XX
DT 07-FEB-2000 (first entry)
XX
DE Complementary DNA consisting of doxA, dnrV and C-terminal dnrU genes.
XX
KW DoxA gene; daunorubicin 14-hydroxylase; daunorubicin; doxorubicin;
KW anticancer agent; dnrV gene; dnrU gene; complementary; ss.
XX
OS Streptomyces peucetius.
XX
PN WO9955829-A2.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US07016.
XX
PR 24-APR-1998; 98US-0065606.
XX
PA (PHAA) PHARMACIA & UPJOHN SPA.
XX
PI Solari AJ, Zanuso G, Filippini S, Torti F, Otten S, Colombo AL;
PI Hutchinson CR;
XX
DR WPI; 2000-023353/02.
XX
PT New DNA containing the gene for daunorubicin-14-hydroxylase and genes
PT for resistance to anthracyclines, used to prepare the anticancer agent

PT doxorubicin -
XX
PS Disclosure; Page 16-19; 35pp; English.
XX
CC The invention relates to a DNA comprising the doxA gene encoding
CC daunorubicin 14-hydroxylase and at least 1 gene (II) that confers
CC daunorubicin (DA) and doxorubicin (DO) resistance. Cells expressing the
CC DNA are used to convert DA to DO. DO is a well known anticancer agent.
CC (II) imparts a high degree of resistance to DO in the host cells and so
CC provides a more efficient conversion process. The present sequence
CC represents a DNA consisting of the doxA, dnrV and the C-terminal part of
CC the dnrU (deldnrU) genes. This sequence is the complementary strand to
CC the coding sequence.
XX
SQ Sequence 2870 BP; 386 A; 1010 C; 1093 G; 381 T; 0 other;

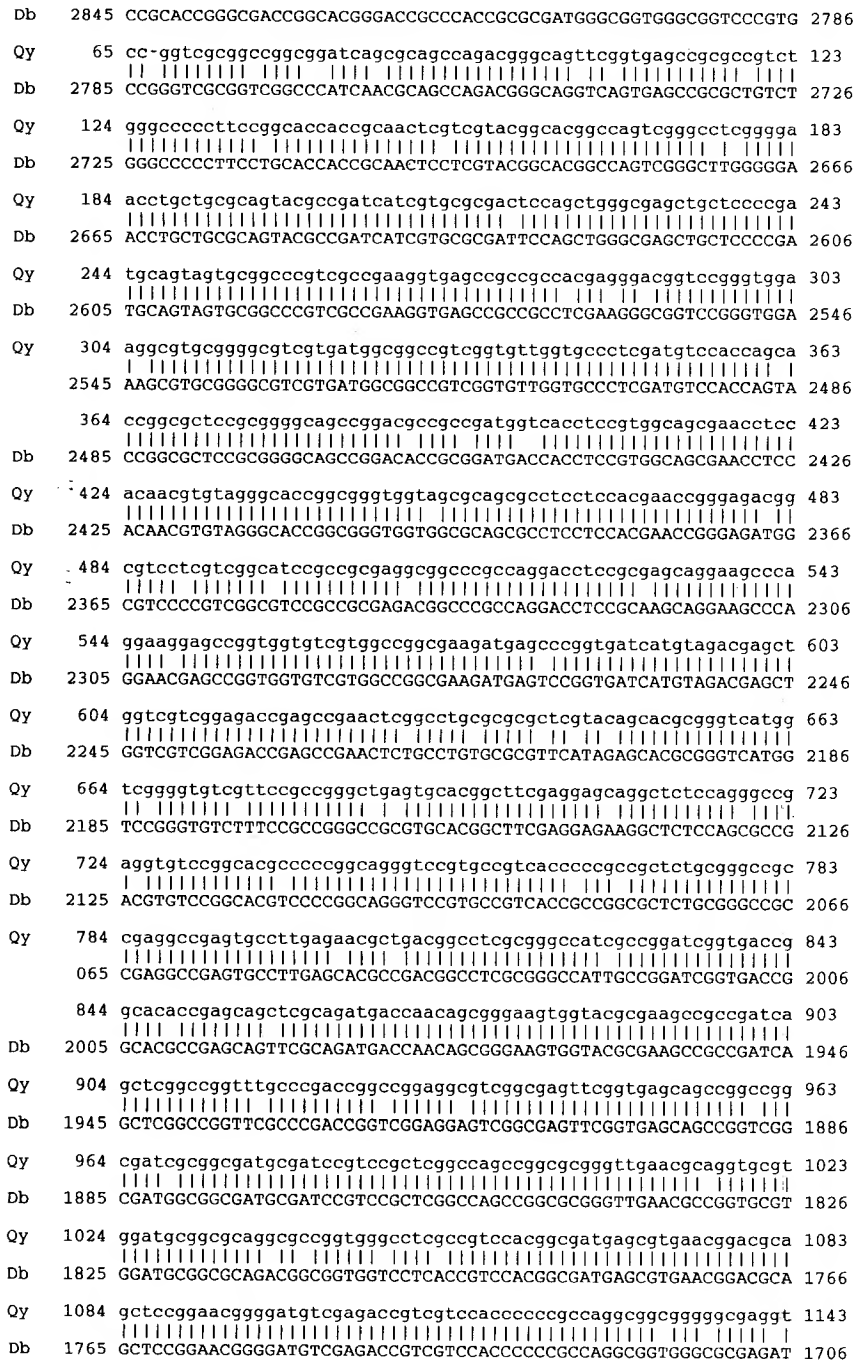
Query Match 100.0%; Score 2870; DB 21; Length 2870;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ggatccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcc 60
Db       1 ggatccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcc 60

Qy      61 cgtgccggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccg 120
Db      61 cgtgccggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccg 120

Qy     121 tctgggccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgg 180
Db     121 tctgggccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgg 180

Qy     181 ggaacctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccc 240
Db     181 ggaacctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccc 240

Qy     241 cgatgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggt 300
Db     241 cgatgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggt 300

Qy     301 ggaaggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccacca 360
Db     301 ggaaggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccacca 360

Qy     361 gcaccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacc 420
Db     361 gcaccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacc 420

Qy     421 tccacaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggaga 480
Db     421 tccacaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggaga 480

Qy     481 cggcgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagc 540
Db     481 cggcgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagc 540

Qy     541 ccaggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacga 600
Db     541 ccaggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacga 600

Qy     601 gctggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtca 660
Db     601 gctggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtca 660

Qy     661 tggtcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccaggg 720
Db     661 tggtcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccaggg 720

Qy     721 ccgaggtgtccggcacgccccggcagggtccgtgccgtcaccccgccgctctgcgggc 780
Db     721 ccgaggtgtccggcacgccccggcagggtccgtgccgtcaccccgccgctctgcgggc 780

Qy     781 cgccgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtga 840
Db     781 cgccgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtga 840

Qy     841 ccggcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccga 900
Db     841 ccggcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccga 900

Qy     901 tcagctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggc 960
Db     901 tcagctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggc 960

Qy     961 cggcgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtg 1020
Db     961 cggcgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtg 1020

Qy    1021 cgtggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggac 1080
Db    1021 cgtggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggac 1080

Qy    1081 gcagctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcga 1140
Db    1081 gcagctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcga 1140

Qy    1141 ggtcggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcgg 1200
Db    1141 ggtcggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcgg 1200

Qy    1201 tgatgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctccc 1260
Db    1201 tgatgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctccc 1260

Qy    1261 ggaaggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggt 1320
Db    1261 ggaaggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggt 1320

Qy    1321 cgacggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtaccc 1380
Db    1321 cgacggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtaccc 1380

Qy    1381 ctcccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccg 1440
Db    1381 ctcccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccg 1440

Qy    1441 ccccctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccg 1500
Db    1441 ccccctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccg 1500

Qy    1501 gaccgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgcc 1560
Db    1501 gaccgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgcc 1560

Qy    1561 ggtgacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtc 1620
Db    1561 ggtgacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtc 1620

Qy    1621 ggtgacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtc 1680
Db    1621 ggtgacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtc 1680

Qy    1681 ccgggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcc 1740
Db    1681 ccgggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcc 1740

Qy    1741 cgtcacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggg 1800
Db    1741 cgtcacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggg 1800

Qy    1801 gtcggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctc 1860
Db    1801 gtcggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctc 1860

Qy    1861 gccgcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacac 1920
```

```
Db    1861 gccgcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacac 1920

Qy    1921 ggtccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggg 1980
Db    1921 ggtccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggg 1980

Qy    1981 ggtgtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgt 2040
Db    1981 ggtgtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgt 2040

Qy    2041 gtatcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgta 2100
Db    2041 gtatcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgta 2100

      2101 gaagtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccggg 2160
Db    2101 gaagtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccggg 2160

Qy    2161 cgcgaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgc 2220
Db    2161 cgcgaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgc 2220

Qy    2221 ggatgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtc 2280
Db    2221 ggatgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtc 2280

Qy    2281 agtgcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccgg 2340
Db    2281 agtgcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccgg 2340

Qy    2341 cgtcggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtgg 2400
Db    2341 cgtcggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtgg 2400

Qy    2401 tcaactcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcgg 2460
Db    2401 tcaactcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcgg 2460

Qy    2461 agcgcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcggg 2520
Db    2461 agcgcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcggg 2520

Qy    2521 tgcggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcct 2580
      2521 tgcggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcct 2580

      2581 ccgccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagc 2640
Db    2581 ccgccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagc 2640

Qy    2641 ggtgacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacg 2700
Db    2641 ggtgacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacg 2700

Qy    2701 aggtgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagc 2760
Db    2701 aggtgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagc 2760

Qy    2761 ctgcgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgacc 2820
Db    2761 ctgcgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgacc 2820

Qy    2821 agaacatgccgccggcgttgctggccatgacatcgatgcgcgggtaccgg 2870
Db    2821 agaacatgccgccggcgttgctggccatgacatcgatgcgcgggtaccgg 2870
```

```
RESULT   2
AAV01453/c
ID   AAV01453 standard; DNA; 3196 BP.
XX
AC   AAV01453;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Streptomyces daunomycin biosynthesis gene cluster.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy; ds.
XX
OS   Streptomyces sp. strain C5.
XX
FH   Key             Location/Qualifiers
FT   CDS             9..590
FT                   /*tag= a
FT                   /note= "3' end of orf1 coding region"
FT   RBS             663..666
FT                   /*tag= b
FT   CDS             674..1501
FT                   /*tag= c
FT                   /product= daunomcin C-14 hydroxylase
FT                   /note= "orfA"
FT   CDS             1498..2766
FT                   /*tag= d
FT                   /note= "doxA"
FT   RBS             3038..3041
FT                   /*tag= e
FT   CDS             3050..3196
FT                   /*tag= f
FT                   /note= "5' end of dauI coding region"
XX
PN   WO9744439-A2.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08690.
XX
PR   24-MAY-1996;   96US-0653650.
XX
PA   (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI   DeSanti CL,  Dickens ML,  Strohl WA;
XX
DR   WPI; 1998-018495/02.
DR   P-PSDB; AAW36128, AAW36130 AND AAW36133-34.
XX
PT   Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT   - also hydroxylation and oxidation of other anthracycline(s) with
PT   the same enzyme
XX
PS   Disclosure; Fig 3; 59pp; English.
XX
CC   This DNA sequence comprises a 3196 bp KpnI-SstI DNA fragment of
CC   Streptomyces sp. strain C5 containing the doxA gene (see also
CC   AAV01447) that codes for daunomycin C-14 hydroxylase (see AAW36128).
CC   The DNA fragment also includes the 3' end of orf1 (see also
CC   AAW36133), the complete orfA (see AAW36134) and the 5' end of dauI
CC   (see 36130), a putative transcriptional activator.  Daunomycin
CC   C-14 hydroxylase is an enzyme capable of converting daunomycin to
CC   the anticancer agent doxorubicin.  Host cells, especially
CC   Streptomyces host cells, transformed with plasmids that include
CC   the doxA gene can be used in methods for the production of
CC   doxorubicin from daunomycin or for the hydroxylation and oxidation
CC   of other anthracyclines.
XX
SQ   Sequence 3196 BP; 475 A; 1182 C; 1078 G; 461 T; 0 other;
```

```
  Query Match             89.1%;  Score 2558.2;  DB 19;  Length 3196;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 2691;  Conservative    0;  Mismatches  153;  Indels   21;  Gaps    2;

Qy       5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
```


Db 2845 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2786
Qy 65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
Db 2785 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2726
Qy 124 gggccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
Db 2725 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2666
Qy 184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
Db 2665 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2606
Qy 244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
Db 2605 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2546
Qy 304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
Db 2545 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2486
Qy 364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
Db 2485 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2426
Qy 424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
Db 2425 ACAACGTGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2366
Qy 484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
Db 2365 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2306
Qy 544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
Db 2305 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2246
Qy 604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
Db 2245 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2186
Qy 664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
Db 2185 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 2126
Qy 724 aggtgtccggcacgccccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
Db 2125 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 2066
Qy 784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
Db 065 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 2006
Qy 844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
Db 2005 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1946
Qy 904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
Db 1945 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1886
Qy 964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
Db 1885 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1826
Qy 1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
Db 1825 GGATGCGGCGCAGACGGCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1766
Qy 1084 gctccggaacggggatgtcgagaccgtcgtccacccccgccaggcggcgggggcgaggt 1143
Db 1765 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1706
Qy 1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
Db 1705 CGGGGTCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1646
Qy 1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
Db 1645 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 1586
Qy 1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
Db 1585 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1526
Qy 1324 cggccacccgggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtacccctc 1383
Db 1525 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCGTCCGCCGCCGCGGGGTACGCCTC 1466
Qy 1384 ccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccgccc 1443
Db 1465 CCGCAGTTCGACCACCGAGAAGCCAGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCGCCC 1406
Qy 1444 cctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccggac 1503
Db 1405 CCTGGGGGTGTCGACGGGCTCGTTCTCGACGGAGCCGCCGAGTTCGACGGCCCGCCGGAC 1346
Qy 1504 cgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgccggt 1563
Db 1345 CGTCGCGTCGCAGTCGTGCACGGAAAACAGCACGGCCCAGTGGGGCCGTACCGCGCCGGT 1286
Qy 1564 gacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtcggt 1623
Db 1285 CGTGCCCAGTTCCCGGGTGCCGGCGACCGGTGTGTCATGGATGTGCCAAACCGGGTCGGT 1226
Qy 1624 gacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtcccg 1683
Db 1225 GACGCCCTTGCGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCACGGTAGAAGCCACG 1166
Qy 1684 ggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcccgt 1743
Db 1165 GGCGGTCCCGATGTCGTCGGTCACCAGTTCCACCCAGCCGACCGAGCCGGGCACGCCCGT 1106
Qy 1744 cacctccgcgcccctccatgactcccttgcgccagaccgcgaacgcggccccggcggggtc 1803
Db 1105 CACCTCCGCGCCCTCCATGACGCCCTTGCGCCACACCGCGAACGCGGCCCCCGATGGGTC 1046
Qy 1804 ggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctcgcc 1863
Db 1045 GGCGAAGACCGCCATCCGGCCGAGACCGAGGACGTCCATCGGGGTCATGATGACGTCGCC 986
Qy 1864 gcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacacggt 1923
Db 985 ACCCGCCGTCTCGACCCGTTTGGTCAGTGCGTCGGCGTCGTTGGTGGCGAAGTACACGGT 926
Qy 1924 ccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggggct 1983
Db 925 CCAGATGGCCGGCATGCCGTGCTGGTCGTTGCCGGGCCCGTACGGACGGTGGTAGGGCGT 866
Qy 1984 gtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgtgta 2043
Db 865 GTCGATCTGATGGCGGGCGACCGCGGCGACCAGCTTCCCGTCGGAGCTGAACGTCGTGTA 806
Qy 2044 tcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgtagaa 2103
Db 805 TCCCCCGGCACCCGGGTCGCTGACCACGGTCGCGGTCCAGCCGAAGAGGCCGGTGTAGAA 746
Qy 2104 gtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccgggcgc 2163
Db 745 GTCGGCCGAGGCGGCGACATCGGGCGACCCGAGGTCGAACCATGCGGGGGCGCCGGGCGC 686
Qy 2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
Db 685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632

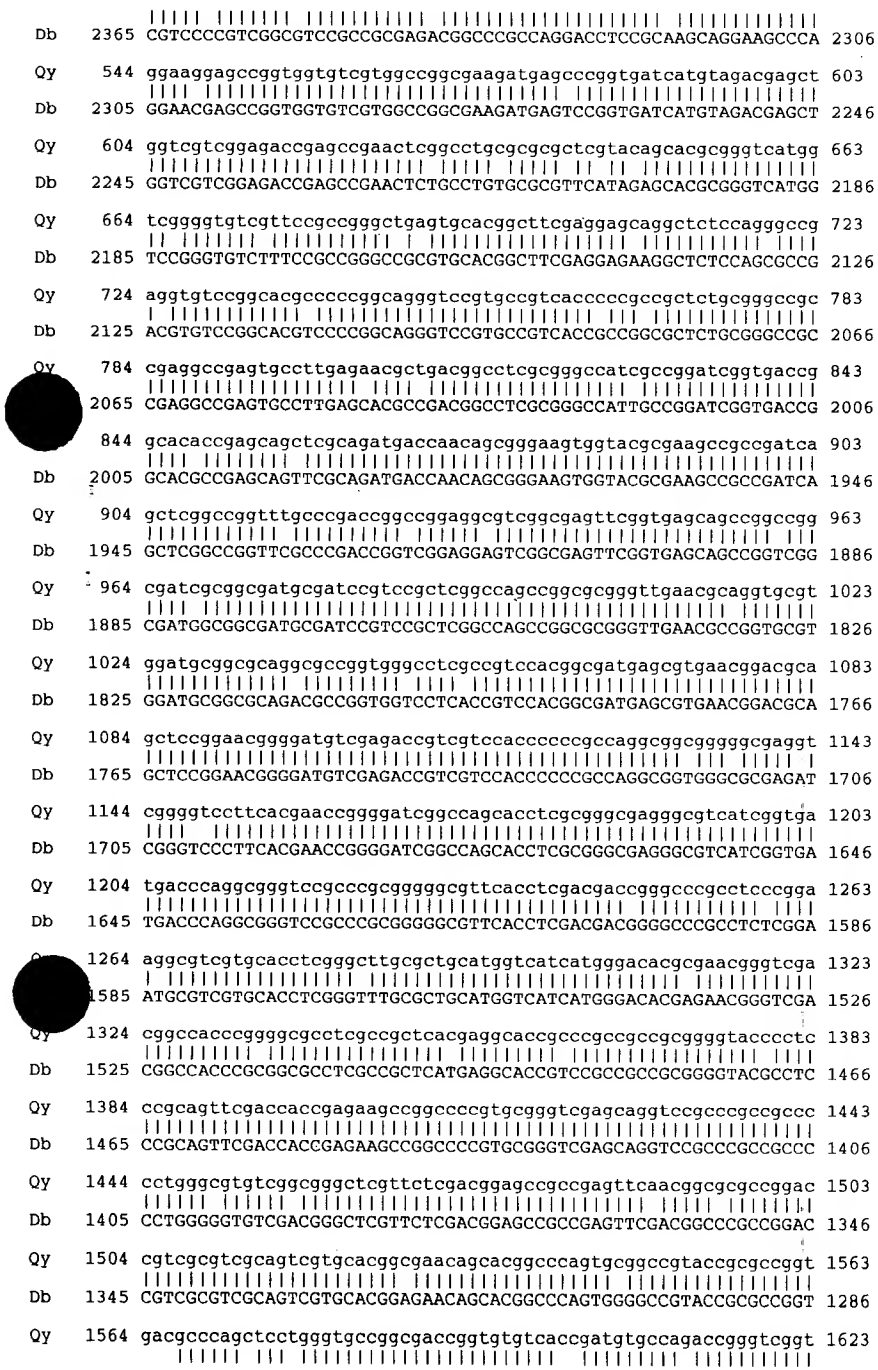
```
Qy   2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
                  |||||||||||||||| |||||||||||| |  ||||||| |||||
Db    633 ----------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGACCCGCCGCGGTCAGC 586

Qy   2284 gcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccggcgt 2343
          |||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||
Db    585 GCGCGGTGTCGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526

Qy   2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAATAGCCGCCCGTGGTCA 466

Qy   2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
          |||||||||||||||| |||||||||||||| |||||||||||||| |||||||| |
       465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406

       464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
          |||||||||||||||||||||||||| ||||||||| |||||||| ||||||||||||||
Db    405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346

Qy   2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
          ||||||| |||||||| |||||||||| |||||||||||||||||||| ||||||||||||
Db    345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286

Qy   2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
            |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226

Qy   2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166

Qy   2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
          ||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||| |
Db    165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGCGCCAGCAGGAAGCCCG 106

Qy   2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
          |||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46

Qy   2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868
          |||||||||||||||||| |||||||||||||||||||||||||
Db     45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1
```

ULT 3
447/c
AAV01447 standard; DNA; 3196 BP.

AC AAV01447;
XX
DT 21-MAY-1998 (first entry)
XX
DE Daunomycin C-14 hydroxylase doxA gene.
XX
KW Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW 13-dihydrocarminomycin; carminomycin; anthracycline;
KW anticancer; cytostatic; cancer; therapy; ds.
XX
OS Streptomyces sp. strain C5.
XX
FH Key Location/Qualifiers
FT CDS 1498..2767
FT /*tag= a
FT mat_peptide 1498..2764
FT /*tag= b
XX
PN WO9744439-A2
XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08690.
XX
PR 24-MAY-1996; 96US-0653650.
XX
PA (OHIS) UNIV OHIO STATE RES FOUND.
XX
PI DeSanti CL, Dickens ML, Strohl WA;
XX
DR WPI; 1998-018495/02.
DR P-PSDB; AAW36128.
XX
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT - also hydroxylation and oxidation of other anthracycline(s) with
PT the same enzyme
XX
PS Claim 11; Page 30-33; 59pp; English.
XX
CC This DNA sequence comprises the doxA gene of Streptomyces sp.
CC strain C5 that codes for daunomycin C-14 hydroxylase (DC14H)
CC (see AAW36128), a P450-like enzyme useful for the production of
CC doxorubicin from daunomycin, and for the hydroxylation and
CC oxidation of other anthracyclines. The gene was identified by
CC sequencing an approximately 8 kb fragment of the daunomycin
CC biosynthesis gene cluster from C5 (see AAV01453). The gene has been
CC incorporated into various vectors, including pANT195 (see AAV01451)
CC and pANT199 (see AAV01452). Claimed methods for production of
CC doxorubicin include: the use of a host microorganism transformed
CC with a plasmid which contains the doxA gene; incubating a culture
CC of Streptomyces sp. strain C5 with daunomycin; and incubating
CC DC14H with daunomycin. The enzyme can also be used in claimed
CC methods of producing 13-dihydrocarminomycin and carminomycin
CC from 13-deoxycarminomycin, and of producing 13-dihydrodaunomycin
CC and daunomycin from 13-deoxydaunomycin. The methods preclude the
CC use of halogens as required in the current chemical methods.
XX
SQ Sequence 3196 BP; 475 A; 1185 C; 1076 G; 460 T; 0 other;

Query Match 88.6%; Score 2542.2; DB 19; Length 3196;
Best Local Similarity 93.6%; Pred. No. 0;
Matches 2681; Conservative 0; Mismatches 163; Indels 21; Gaps 2;

```
Qy      5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
          |||||||||    |||||||||||||||||||||| || ||||||||||||||||||
Db   2845 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2786

Qy     65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
          || ||||||||| ||||  |||| |||||||||||||||||| || |||||||||||| ||||
Db   2785 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2726

Qy    124 gggccccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
          |||||||||||| |||||||||||||| |||||||||||||||||||||||| | |||||
Db   2725 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2666

Qy    184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   2665 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2606

Qy    244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
          |||||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||||
Db   2605 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2546

Qy    304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2545 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2486

Qy    364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
          |||||||||||||||||||||||| |||| ||||  ||||||||||||||||||||||||
Db   2485 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2426

Qy    424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
          |||||| |||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db   2425 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2366

Qy    484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
```


Db 2365 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2306
Qy 544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
Db 2305 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2246
Qy 604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
Db 2245 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2186
Qy 664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
Db 2185 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 2126
Qy 724 aggtgtccggcacgccccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
Db 2125 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 2066
Qy 784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
Db 2065 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 2006
Qy 844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
Db 2005 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1946
Qy 904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
Db 1945 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1886
Qy 964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
Db 1885 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1826
Qy 1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
Db 1825 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1766
Qy 1084 gctccggaacggggatgtcgagaccgtcgtccacccccgccaggcggcgggggcgaggt 1143
Db 1765 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1706
Qy 1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
Db 1705 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1646
Qy 1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
Db 1645 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 1586
Qy 1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
Db 1585 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1526
Qy 1324 cggccacccggggcgcctcgccgctcacgaggcaccgcccgccgccgcggggtaccctc 1383
Db 1525 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCGTCCGCCGCCGCGGGGTACGCCTC 1466
Qy 1384 ccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccgccc 1443
Db 1465 CCGCAGTTCGACCACCGAGAAGCCGGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCGCCC 1406
Qy 1444 cctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccggac 1503
Db 1405 CCTGGGGGTGTCGACGGGCTCGTTCTCGACGGAGCCGCCGAGTTCGACGGCCCGCCGGAC 1346
Qy 1504 cgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgccggt 1563
Db 1345 CGTCGCGTCGCAGTCGTGCACGGAGAACAGCACGGCCCAGTGGGGCCGTACCGCGCCGGT 1286
Qy 1564 gacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtcggt 1623
Db 1285 CGTGCCCAGTTCCCGGGTGCCGGCGACCGGTGTGTCATGGATGTGCCAAACCGGGTCGGT 1226
Qy 1624 gacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtcccg 1683
Db 1225 GACGCCCTTGCGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCACGGTAGAAGCCACG 1166
Qy 1684 ggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcccgt 1743
Db 1165 GGCGGTCCCGATGTCGTCGGTCACCAGTTCCACCCAGCCGACCGAGCCGGGCACGCCCGT 1106
Qy 1744 cacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggggtc 1803
Db 1105 CACCTCCGCGCCCTCCATGACGCCCTTGCGCCACACCGCGAACGCGGCCCCCGATGGGTC 1046
Qy 1804 ggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctcgcc 1863
Db 1045 GGCGAAGACCGCCATCCGGCCGAGACCGAGGACGTCCATCGGGTGCATGATGACGTCGCC 986
Qy 1864 gcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacacggt 1923
Db 985 ACCCGCCGCTTCGACCCGTTTGGTCAGTGCGTCGGCGTCGTTGGTGGCGAAGTACACGGT 926
Qy 1924 ccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggggct 1983
Db 925 CCAGATGGCCGGCATGCCGTGCTGGTGCTTGCCGGGCCCGTACGGACGGTGGTAGGGCGT 866
Qy 1984 gtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgtgta 2043
Db 865 GTCGATCTGATGGCGGGCGACCGCGGCGACAGGCTTCCCGTCGGAGCTGAAAGTAGTGTA 806
Qy 2044 tcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgtagaa 2103
Db 805 TCCCCCGGCACCCGGGTCGCTGACCACGGTCGCGGTCCACGCGAAGAGGCCGGTGTAGAA 746
Qy 2104 gtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccgggcgc 2163
Db 745 GTCGGCCGAGGCGGCGACATCGGGCGACCCGAGGTCGAACCATGCGGGGGCGCCGGGCGC 686
Qy 2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
Db 685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632
Qy 2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
Db 633 ----------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGGCCCGCCGCGGTCAGC 586
Qy 2284 gcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccggcgt 2343
Db 585 GCGCGGTGTGGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526
Qy 2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
Db 525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCA 466
Qy 2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
Db 465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406
Qy 2464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
Db 405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346
Qy 2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
Db 345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286
Qy 2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
Db 285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226
Qy 2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
Db 225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166

```
Qy   2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCCG 106

Qy   2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
          |||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46

Qy   2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868
          ||||||||||||||||| ||||||||||||||||||||||||||
Db     45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1
```

LT 4
617/c
AAT33617 standard; DNA; 1269 BP.

AC AAT33617;
XX
DT 03-JAN-1997 (first entry)
XX
DE DxrA gene encoding daunorubicin 14-hydroxylase.
XX
KW dxrA; doxorubicin; bioconversion; anthracycline; antitumour therapy;
KW daunorubicin 14-hydroxylase; ds.
XX
OS Streptomyces peucetius strain 29050.
XX
PN WO9627014-A1.
XX
PD 06-SEP-1996.
XX
PF 20-FEB-1996; 96WO-EP00692.
XX
PR 27-FEB-1995; 95US-0396218.
XX
PA (PHAA) PHARMACIA SPA.
XX
PI Breme U, Colombo AL, Hutchinson CR, Otten S, Scotti C;
PI Solari INVENTI A, Inventi AS;
XX
DR WPI; 1996-412780/41.
DR P-PSDB; AAW00729.
XX
New DNA encoding daunorubicin 14-hydroxylase - useful for converting daunorubicin to the antitumour agent doxorubicin

Claim 6; Page 15-18; 31pp; English.
XX
CC The present sequence is the drxA gene coding sequence for daunorubicin
CC 14-hydroxylase, isolated from Streptomyces peucetius strain 29050. The
CC enzyme converts daunorubicin to doxorubicin, an anthracycline used widely
CC as an antitumour agent. A vector contg. the drxA gene could be used to
CC transform a suitable host cell which produces daunorubicin, in order to
CC enhance bioconversion of daunorubicin to doxorubicin. Bioconversion can
CC be carried out either by using directly the free or immobilised
CC transformed cells or by isolating the enzyme, which can be used in the
CC free form or immobilised to resins, glass, cellulose or similar
CC substances by ionic or covalent bonds, or grafted to fibres permeable to
CC the substrate or insolubilised by cross-linkage.
XX
SQ Sequence 1269 BP; 164 A; 484 C; 442 G; 179 T; 0 other;

Query Match 44.2%; Score 1269; DB 17; Length 1269;
Best Local Similarity 100.0%; Pred. No. 2.9e-167;
Matches 1269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     83 tcagcgcagccagacgggcagttcggtgagccgcgccgtctgggccccttccggcacca 142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 TCAGCGCAGCCAGACGGGCAGTTCGGTGAGCCGCGCCGTCTGGGCCCCCTTCCGGCACCA 1210

Qy    143 ccgcaactcgtcgtacggcacggccagtcgggcctcggggaacctgctgcgcagtacgcc 202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 CCGCAACTCGTCGTACGGCACGGCCAGTCGGGCCTCGGGGAACCTGCTGCGCAGTACGCC 1150

Qy    203 gatcatcgtgcgcgactccagctgggcgagctgctccccgatgcagtagtgcggcccgtc 262
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 GATCATCGTGCGCGACTCCAGCTGGGCGAGCTGCTCCCCGATGCAGTAGTGCGGCCCGTC 1090

Qy    263 gccgaaggtgagccgccgccacgagggacggtccgggtggaaggcgtgcggggcgtcgtg 322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 GCCGAAGGTGAGCCGCCGCCACGAGGGACGGTCCGGGTGGAAGGCGTGCGGGGCGTCGTG 1030

Qy    323 atggcggccgtcggtgttggtgccctcgatgtccaccagcaccggcgctccgcggggcag 382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 ATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCCGCGGGGCAG 970

Qy    383 ccggacgccgccgatggtcacctccgtggcagcgaacctccacaacgtgtagggcaccgg 442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 CCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACCTCCACAACGTGTAGGGCACCGG 910

Qy    443 cgggtggtagcgcagcgcctcctccacgaaccgggagacggcgtcctcgtcggcatccgc 502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 CGGGTGGTAGCGCAGCGCCTCCTCCACGAACCGGGAGACGGCGTCCTCGTCGGCATCCGC 850

Qy    503 cgcgaggcggcccgccaggacctccgcgagcaggaagcccaggaaggagccggtggtgtc 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 CGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGGAGCCGGTGGTGTC 790

Qy    563 gtggccggcgaagatgagcccggtgatcatgtagacgagctggtcgtcggagaccgagcc 622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 GTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACGAGCTGGTCGTCGGAGACCGAGCC 730

Qy    623 gaactcggcctgcgcgcgctcgtacagcacgcgggtcatggtcggggtgtcgttccgccg 682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCATGGTCGGGGTGTCGTTCCGCCG 670

Qy    683 ggctgagtgcacggcttcgaggagcaggctctccagggccgaggtgtccggcacgccccc 742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GGCTGAGTGCACGGCTTCGAGGAGCAGGCTCTCCAGGGCCGAGGTGTCCGGCACGCCCCC 610

Qy    743 ggcagggtccgtgccgtcaccccgccgctctgcgggccgccgaggccgagtgccttgag 802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 GGCAGGGTCCGTGCCGTCACCCCCGCCGCTCTGCGGGCCGCCGAGGCCGAGTGCCTTGAG 550

Qy    803 aacgctgacggcctcgcgggccatcgccggatcggtgaccggcacaccgagcagctcgca 862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 AACGCTGACGGCCTCGCGGGCCATCGCCGGATCGGTGACCGGCACACCGAGCAGCTCGCA 490

Qy    863 gatgaccaacagcgggaagtggtacgcgaagccgccgatcagctcggccggtttgcccga 922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 GATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCAGCTCGGCCGGTTTGCCCGA 430

Qy    923 ccggccggaggcgtcggcgagttcggtgagcagccggccggcgatcgcggcgatgcgatc 982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 CCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCCGGCGATCGCGGCGATGCGATC 370

Qy    983 cgtccgctcggccagccggcgcgggttgaacgcaggtgcgtggatgcggcgcaggcgccg 1042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTGCGTGGATGCGGCGCAGGCGCCG 310

Qy   1043 gtgggcctcgccgtccacggcgatgagcgtgaacggacgcagctccggaacggggatgtc 1102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 GTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGACGCAGCTCCGGAACGGGGATGTC 250

Qy   1103 gagaccgtcgtccacccccccgccaggcggcgggggcgaggtcgggggtccttcacgaaccg 1162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 GAGACCGTCGTCCACCCCCCCGCCAGGCGGCGGGGGCGAGGTCGGGGTCCTTCACGAACCG 190

Qy   1163 gggatcggccagcacctcgcgggcgagggcgtcatcggtgatgacccaggcgggtccgcc 1222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 GGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGATGACCCAGGCGGGTCCGCC 130

Qy   1223 cgcgggggcgttcacctcgacgaccgggcccgcctcccggaaggcgtcgtgcacctcggg 1282
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 CGCGGGGGCGTTCACCTCGACGACCGGGCCCGCCTCCCGGAAGGCGTCGTGCACCTCGGG 70

Qy    1283 cttgcgctgcatggtcatcatgggacacgcgaacgggtcgacggccacccggggcgcctc 1342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 CTTGCGCTGCATGGTCATCATGGGACACGCGAACGGGTCGACGGCCACCCGGGGCGCCTC 10

Qy    1343 gccgctcac 1351
           |||||||||
Db       9 GCCGCTCAC 1


RESULT   5
AAV01451/c
ID   AAV01451 standard; DNA; 3013 BP.
XX
AC   AAV01451;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Plasmid pANT195 including the doxA gene.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy; plasmid pANT195; ds.
XX
OS   Streptomyces sp. strain C5.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (99..1034)
FT                   /*tag=  a
FT                   /note= "snpR activator gene"
FT   -10_signal      complement (1043..1047)
FT                   /*tag=  b
FT   -35_signal      complement (1067..1071)
FT                   /*tag=  c
FT   protein_bind    1193..1209
FT                   /*tag=  d
FT                   /note= "proposed SnpR binding site"
FT   -35_signal      1230..1235
FT                   /*tag=  e
FT   -10_signal      1252..1257
FT                   /*tag=  f
FT   RBS             1262..1265
FT                   /*tag=  g
FT   RBS             1302..1305
FT                   /*tag=  h
FT                   /note= "engineered RBS site"
FT   CDS             1315..2583
FT                   /*tag=  i
FT                   /note= "doxA gene"
FT   RBS             2855..2858
FT                   /*tag=  j
FT   CDS             2867..3013
FT                   /*tag=  k
FT                   /note= "5' end of dauI gene"
XX
PN   WO9744439-A2.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08690.
XX
PR   24-MAY-1996;  96US-0653650.
XX
PA   (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI   DeSanti CL,  Dickens ML,  Strohl WA;
XX
DR   WPI; 1998-018495/02.
DR   P-PSDB; AAW36128-30.
XX
PT   Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT   - also hydroxylation and oxidation of other anthracycline(s) with
PT   the same enzyme
XX
PS   Disclosure; Fig 6; 59pp; English.
XX
CC   This DNA sequence comprises a fragment of plasmid pANT195
CC   containing the Streptomyces snpR activator gene, the SnpR-
CC   activated promoter and a 5' end-modified (added ribosome binding
CC   site) doxA gene (see also AAV01447) derived from Streptomyces sp.
CC   strain C5.  The doxA gene codes for daunomycin C-14 hydroxylase
CC   (see AAW36128), an enzyme capable of converting daunomycin to
CC   the anticancer agent doxorubicin.  Host cells, especially
CC   Streptomyces host cells, transformed with pANT195 can be used
CC   in methods for the production of doxorubicin from daunomycin or for
CC   the hydroxylation and oxidation of other anthracyclines.
XX
SQ   Sequence 3013 BP; 448 A; 1074 C; 1066 G; 425 T; 0 other;


  Query Match             42.7%;  Score 1224.8;  DB 19;  Length 3013;
  Best Local Similarity   94.3%;  Pred. No. 3.2e-161;
  Matches 1282;  Conservative    0;  Mismatches   77;  Indels    1;  Gaps    1;

Qy       5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
           |||||||||||    |||||||||||||||||||||||| || ||||||||||||||||||
Db    2662 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2603

Qy      65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
           || |||||||| ||||  |||| |||||||||||||||||| || ||||||||||| ||||
Db    2602 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2543

Qy     124 gggccccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
           |||||||||||| |||||||||||||| |||||||||||||||||||||| | |||||
Db    2542 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2483

Qy     184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    2482 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2423

Qy     244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
           |||||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||
Db    2422 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2363

Qy     304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    2362 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2303

Qy     364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
           |||||||||||||||||||||||| |||| ||||  ||||||||||||||||||||||||
Db    2302 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2243

Qy     424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    2242 ACAACGTGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2183

Qy     484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
           ||||| ||||||| |||||||||||| ||||||||||||||||||||| ||||||||||||
Db    2182 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2123

Qy     544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
           |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    2122 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2063

Qy     604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
           |||||||||||||||||||||||| ||||| ||||| || || ||||||||||||||||
Db    2062 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2003

Qy     664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
           || ||||||| |||||||||| | ||||||||||||||||| |||||||||| ||||
Db    2002 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 1943

Qy     724 aggtgtccggcacgccccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
           | |||||||||||| |||||||||||||||||||||||| ||| ||||||||||||||||
```

```
Db   1942 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 1883

Qy    784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
          |||||||||||||||||| |||| |||||||||||||||| ||||||||||||||||
Db   1882 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 1823

Qy    844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
          |||| |||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1822 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1763

Qy    904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
          |||||||||||| |||||||||| |||||| |||||||||||||||||||||||||| |||
Db   1762 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1703

Qy    964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
          |||| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   1702 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1643

Qy   1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
          |||||||||||| || |||||| |||| ||||||||||||||||||||||||||||||
Db   1642 GGATGCGGCGCAGACGGCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1583

Qy   1084 gctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcgaggt 1143
          |||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |
Db   1582 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1523

Qy   1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1522 CGGGGTCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1463

Qy   1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||
Db   1462 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 1403

Qy   1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
          | |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db   1402 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1343

Qy   1324 cggccacccggggcgcctcgccgctcacgaggcaccgccc 1363
          |||||||||| |||||||||||||||| |||||||| | |
Db   1342 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCCCTC 1303


RESULT   6
AAV01452/c
     AAV01452 standard; DNA; 2081 BP.

     AAV01452;

     21-MAY-1998  (first entry)
XX
DE   Modified doxA gene encoding daunomycin C-14 hydroxylase.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy; plasmid pANT199; ds.
XX
OS   Chimeric - Streptomyces sp. strain C5.
OS   Chimeric - synthetic.
XX
FH   Key             Location/Qualifiers
FT   -35_signal       3..8
FT                    /*tag=  a
FT   -10_signal       26..31
FT                    /*tag=  b
FT   misc_signal      38..58
FT                    /*tag=  c
FT                    /note= "lac operator"
FT   misc_signal      74..142
FT                    /*tag=  d
FT                    /note= "rrnB antiterminator"
FT   enhancer         159..168
FT                    /*tag=  e
FT                    /note= "g10 translational enhancer"
FT   RBS              184..188
FT                    /*tag=  f
FT   CDS              197..223
FT                    /*tag=  g
FT                    /note= "mini-cistron"
FT   CDS              227..1651
FT                    /*tag=  h
FT                    /note= "modified doxA gene; the native doxA
FT                           coding sequence comprises nucleotides
FT                           383..1651"
XX
PN   WO9744439-A2.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08690.
XX
PR   24-MAY-1996;   96US-0653650.
XX
PA   (OHIS ) UNIV OHIO STATE RES FOUND.
XX
PI   DeSanti CL,  Dickens ML,  Strohl WA;
XX
DR   WPI; 1998-018495/02.
DR   P-PSDB; AAW36132.
XX
PT   Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT   - also hydroxylation and oxidation of other anthracycline(s) with
PT   the same enzyme
XX
PS   Disclosure; Fig 9; 59pp; English.
XX
CC   This DNA sequence comprises a fragment of plasmid pANT199 in
CC   which the Streptomyces sp. strain C5 doxA gene (see AAV01447) is
CC   translationally fused with a leader sequence encoding 6
CC   histidine residues so that the fusion protein can be affinity
CC   purified on a nickel-agarose gel.  The doxA gene codes for
CC   daunomycin C-14 hydroxylase (see AAW36128), an enzyme capable of
CC   converting daunomycin to the anticancer agent doxorubicin.  Host
CC   cells, especially Streptomyces host cells, transformed with
CC   pANT199 can be used in methods for the production of doxorubicin
CC   from daunomycin or for the hydroxylation and oxidation of other
CC   anthracyclines.
XX
SQ   Sequence 2081 BP; 380 A; 688 C; 657 G; 356 T; 0 other;


  Query Match             42.6%;  Score 1221.6;  DB 19;  Length 2081;
  Best Local Similarity   94.1%;  Pred. No. 9.3e-161;
  Matches 1280;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy       5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
           ||||||||||    |||||||||||||||||||||||| || ||||||||||||||||||
Db    1730 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 1671

Qy      65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
           || |||||||| ||||  |||| |||||||||||||||| || |||||||||| ||||
Db    1670 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 1611

Qy     124 gggcccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
           |||||||||||| |||||||||||||| |||||||||||||||||||||| | |||||
Db    1610 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 1551

Qy     184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccgá 243
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    1550 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 1491

Qy     244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
           |||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||||
Db    1490 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 1431

Qy     304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
```

```
Db    1430 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 1371

Qy     364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
Db    1370 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 1311

Qy     424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
Db    1310 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 1251

Qy     484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
Db    1250 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 1191

Qy     544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
Db    1190 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 1131

Qy     604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
Db    1130 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 1071

Qy     664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
Db    1070 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 1011

Qy     724 aggtgtccggcacgcccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
Db    1010 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 951

Qy     784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
Db     950 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 891

Qy     844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
Db     890 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 831

Qy     904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
Db     830 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 771

Qy     964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
Db     770 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 711

Qy    1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
Db     710 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 651

Qy    1084 gctccggaacggggatgtcgagaccgtcgtccacccccgccaggcggcgggggcgaggt 1143
Db     650 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 591

Qy    1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
Db     590 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 531

Qy    1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
Db     530 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 471

Qy    1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
Db     470 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 411

Qy    1324 cggccacccggggcgcctcgccgctcacgaggcaccgccc 1363
Db     410 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCCCTC 371

RESULT   7
```

```
AAV34001/c
ID   AAV34001 standard; DNA; 1569 BP.
XX
AC   AAV34001;
XX
DT   25-JAN-1999  (first entry)
XX
DE   DNA fragment containing S. peucetius dnrU gene.
XX
KW   dnrU; dnrX; recombinant; microorganism; doxorubicin; daunorubicin;
KW   metabolism; baumycin; secretion; anti-tumour compound; ss.
XX
OS   Streptococcus peucetius.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..255
FT                   /*tag=  a
FT                   /product= "pdsG gene product"
FT                   /transl_except= (pos:34..36, aa:Leu)
FT   CDS             252..1115
FT                   /*tag=  b
FT                   /product= "dnrU gene product"
FT                   /transl_except= (pos:426..428, aa:Met)
FT                   /transl_except= (pos:621..623, aa:Met)
FT                   /transl_except= (pos:624..626, aa:Asn)
FT                   /transl_except= (pos:750..752, aa:Asn)
FT                   /transl_except= (pos:801..803, aa:Asn)
FT                   /transl_except= (pos:855..857, aa:Asn)
FT   CDS             1219..1569
FT                   /*tag=  c
FT                   /product= "partial dnrV gene product"
FT                   /transl_except= (pos:1255..1257, aa:Asp)
FT                   /transl_except= (pos:1258..1260, aa:Gly)
FT                   /transl_except= (pos:1420..1422, aa:Asn)
FT                   /transl_except= (pos:1450..1452, aa:Asn)
FT                   /transl_except= (pos:1456..1458, aa:Asn)
FT                   /transl_except= (pos:1480..1482, aa:Pro)
FT                   /note= "no stop codon is given at the 5' end of
FT                       the sequence"
XX
PN   WO9839458-A1.
XX
PD   11-SEP-1998.
XX
PF   05-MAR-1998;   98WO-US03938.
XX
PR   28-JUL-1997;   97US-0901306.
PR   06-MAR-1997;   97US-0812412.
XX
PA   (PHAA ) PHARMACIA & UPJOHN SPA.
XX
PI   Colombo AL,  Filippini S,  Fonstein L,  Htutchinson RC;
PI   Lomovskaya N;
XX
DR   WPI; 1998-506367/43.
DR   P-PSDB; AAW68558, AAW68559, AAW68560.
XX
PT   Production of doxorubicin and daunorubicin - using a recombinant
PT   daunorubicin producing microorganism in which at least one dnrU or
PT   dnrX metabolism gene has been inactivated
XX
PS   Disclosure; Fig 2A-B; 48pp; English.
XX
CC   This sequence represents an SstI-AatII restriction fragment from the
CC   Streptomyces peucetius genome, which contains the dnrU gene.  The
CC   invention relates to a recombinant microorganism which produces
CC   doxorubicin (DoR), in which at least 1 daunorubicin (DaR) metabolism
CC   gene selected from dnrU and dnrX, has been inactivated. The microorganism
CC   is preferably S. peucetius WMH1658, WMH1654 and WMH1662.  The function
CC   of the DaR gene is inactivated by insertion of a gene (preferably
CC   neomycin/kanamycin resistance gene aphII) into the daunorubicin
CC   metabolism gene.  The blocking of the function of at least one gene
CC   of DaR metabolism can increase DaR and DoR production levels and cause
```

CC the disappearance of baumycin-like products resulting in DaR and DoR
CC secretion directly into the culture medium. The products can be used
CC in the production of anti-tumour compounds.
XX
SQ Sequence 1569 BP; 234 A; 587 C; 545 G; 203 T; 0 other;

Query Match 36.4%; Score 1044.4; DB 19; Length 1569;
Best Local Similarity 99.9%; Pred. No. 2.8e-136;
Matches 1045; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   1825 gaggccgaggacgtccatcggagtcatgatgacctcgccgcccgccgtctcgacccgctt 1884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1569 GAGGCCGAGGACGTCCATCGGAGTCATGATGACCTCGCCGCCCGCCGTCTCGACCCGCTT 1510

Qy   1885 ggtcagtgcgtcggcgtcgtcggtggcgaagtacacggtccagatggccggcatgccgtg 1944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509 GGTCAGTGCGTCGGCGTCGTCGGTGGCGAAGTACACGGTCCAGATGGCCGGCATGCCGTG 1450

Qy   1945 ctggtcgttcccgggcccgtacggccggtggtaggggggtgtcgatctggtggcgggcgac 2004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 CTGGTCGTTCCCGGGCCCGTACGGCCGGTGGTAGGGGGTGTCGATCTGGTGGCGGGCGAC 1390

Qy   2005 cgcggcgaccagcttcccgtcggagctgaacgtcgtgtatcccccggcgcccgggtcgct 2064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 CGCGGCGACCAGCTTCCCGTCGGAGCTGAACGTCGTGTATCCCCCGGCGCCCGGGTCGCT 1330

Qy   2065 gaccacggtggcggtccagccgaacaggccggtgtagaagtcggccgaggcggcgacatc 2124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GACCACGGTGGCGGTCCAGCCGAACAGGCCGGTGTAGAAGTCGGCCGAGGCGGCGACATC 1270

Qy   2125 gggcgaaccgaggtcgaaccatgcgggggcgccgggcgcgaacctggtcacgaatcgttc 2184
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 GGGCGAACCGAGCTCGAACCATGCGGGGGCGCCGGGCGCGAACCTGGTCACGAATCGTTC 1210

Qy   2185 ctttcgatggatcggcacacgagcgtctgcgctcgcggatgagacggacatctcgcggat 2244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 CTTTCGATGGATCGGCACACGAGCGTCTGCGCTCGCGGATGAGACGGACATCTCGCGGAT 1150

Qy   2245 gagacggacatgcgggcggggcgggccgccgccgtcagtgcgcggtgtcgccgacggcgg 2304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 GAGACGGACATGCGGGCGGGGCGGGCCGCCGCCGTCAGTGCGCGGTGTCGCCGACGGCGG 1090

Qy   2305 ccgcgccggcctcccagagcttcgccgcgaggccggcgtcggcggtcgggccgctcaccg 2364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CCGCGCCGGCCTCCCAGAGCTTCGCCGCGAGGCCGGCGTCGGCGGTCGGGCCGCTCACCG 1030

Qy   2365 gggacagccgccggtcgctgtagtagccgcccgtggtcaactcctcggccggcgcggacg 2424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 GGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCAACTCCTCGGCCGGCGCGGACG 970

Qy   2425 ccagccacacgagggtgtcggcgcccttcgccgcggagcgcaggaaggggttgaaccgga 2484
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 CCAGCCACACGAGGGTGTCGGCGCCCTTCGCCGCGGAGCGCAGGAAGGGGTTGAACCGGA 910

Qy   2485 agtaggacgaggcgaccgtgccccgtccgatgcgggtgcggacctcaccggggtgatagc 2544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 AGTAGGACGAGGCGACCGTGCCCCGTCCGATGCGGGTGCGGACCTCACCGGGGTGATAGC 850

Qy   2545 tgaccgccagcacgtccggccagcgcctggcggcctccgccgcggtcatgatgttggcct 2604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 TGACCGCCAGCACGTCCGGCCAGCGCCTGGCGGCCTCCGCCGCGGTCATGATGTTGGCCT 790

Qy   2605 gtttggacgtgccgtacgcctggccggcgctgtagcggtgacggtcgccgttgaggtcgt 2664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 GTTTGGACGTGCCGTACGCCTGGCCGGCGCTGTAGCGGTGACGGTCGCCGTTGAGGTCGT 730

Qy   2665 ccgggtcgatccggccctgggtgtacgcgtcggacgaggtgaggatcagccgcccgcccg 2724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 CCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGGTGAGGATCAGCCGCCCGCCCG 670

Qy   2725 cgagccgctcccgcagcagccgtgccagcaggaagcctgcgaggtgattgacctggatgg 2784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 CGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCTGCGAGGTGATTGACCTGGATGG 610

Qy   2785 tggcctcgaacccgtcctgggtcgtggtgcgcgaccagaacatgccgccggcgttgctgg 2844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 TGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGAACATGCCGCCGGCGTTGCTGG 550

Qy   2845 ccatgacatcgatgcgcgggtaccgg 2870
          |||||||||||||||||||||||||
Db    549 CCATGACATCGATGCGCGGGTACCGG 524
```

Search completed: June 11, 2002, 20:06:24
Job time: 7631 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 16:52:33 ; Search time 55.8 Seconds
(without alignments)
12633.838 Million cell updates/sec

Title: US-09-673-254-1
Perfect score: 2870
Sequence: 1 ggatccgcaccgggtacacg.....catcgatgcgcgggtaccgg 2870

[redacted]ng table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

5962,293 = Nos. 1, 5 & 7
5,976,830 = Nos. 2, 6, 8
5,695,966 = No. 3
5,786,190 = No. 4

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

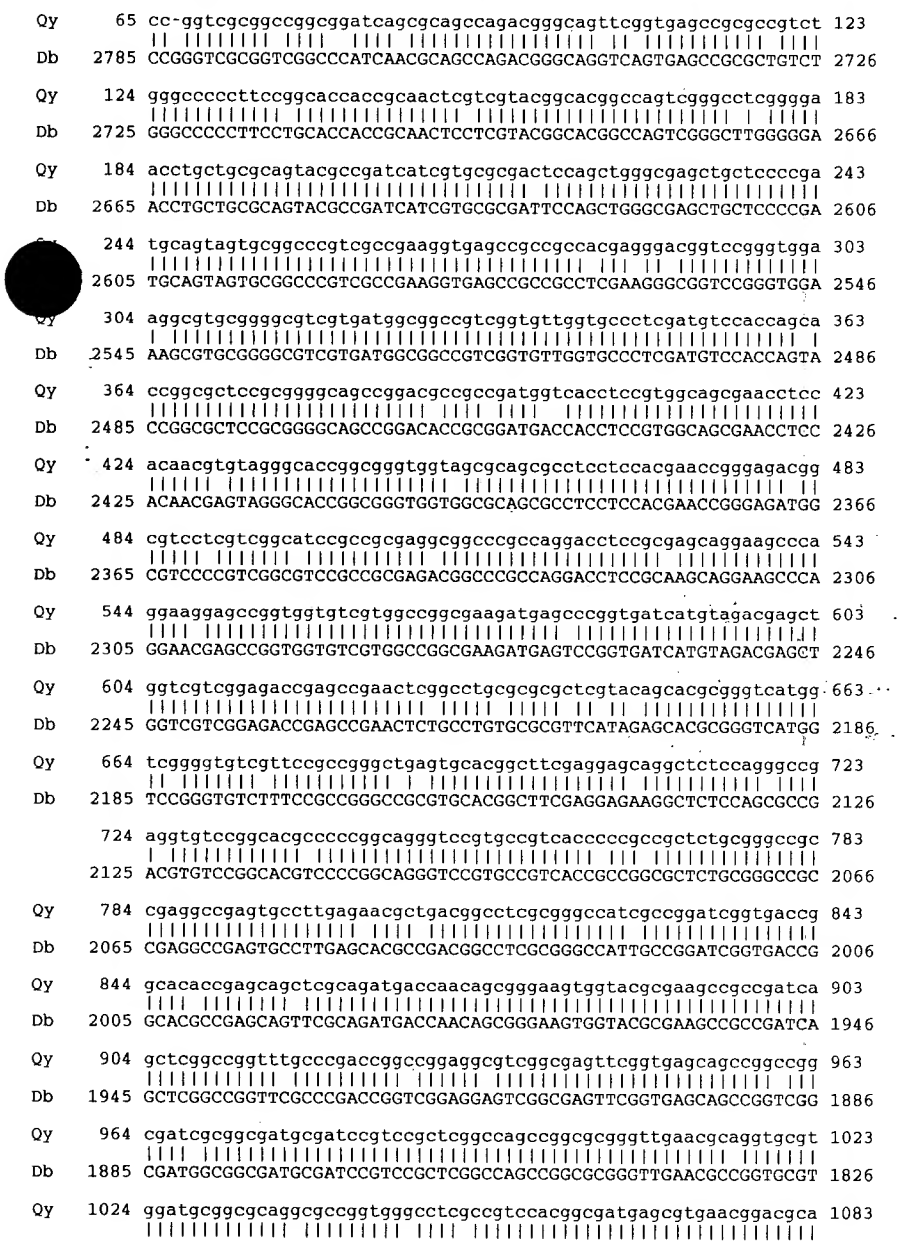
Result No.		Score	% Query Match	Length	DB	ID	Description
1		2542.2	88.6	3196	2	US-09-096-982-4	Sequence 4, Appli
2		2542.2	88.6	3196	2	US-08-653-650A-4	Sequence 4, Appli
3		1269	44.2	1269	1	US-08-396-218-1	Sequence 1, Appli
4		1269	44.2	1269	1	US-08-760-116-1	Sequence 1, Appli
c	5	1221.6	42.6	2081	2	US-09-096-982-7	Sequence 7, Appli
c	6	1221.6	42.6	2081	2	US-08-653-650A-7	Sequence 7, Appli
c	7	1102.2	38.4	3013	2	US-09-096-982-6	Sequence 6, Appli
c	8	1102.2	38.4	3013	2	US-08-653-650A-6	Sequence 6, Appli
c	9	1046	36.4	1569	2	US-08-901-306-3	Sequence 3, Appli
c	10	1044.4	36.4	1569	4	US-09-180-271-3	Sequence 3, Appli
c	11	592	20.6	864	2	US-08-901-306-1	Sequence 1, Appli
c	12	592	20.6	864	4	US-09-180-271-1	Sequence 1, Appli
c	13	137.4	4.8	4257	2	US-08-690-473-1	Sequence 1, Appli
c	14	137.4	4.8	4257	4	US-09-259-821A-1	Sequence 1, Appli
c	15	137.4	4.8	4257	4	US-08-843-659-1	Sequence 1, Appli
c	16	137.4	4.8	5970	3	US-09-320-878-21	Sequence 21, Appl
c	17	135.4	4.7	1251	4	US-09-105-537-38	Sequence 38, Appl
	18	134.2	4.7	12001	1	US-08-458-568A-11	Sequence 11, Appl
c	19	133.2	4.6	4496	4	US-08-765-907A-6	Sequence 6, Appli
c	20	130.6	4.6	44377	2	US-08-804-227C-7	Sequence 7, Appli
c	21	130.6	4.6	44377	2	US-08-804-198-1	Sequence 1, Appli
c	22	128.6	4.5	1735	1	US-08-102-863-10	Sequence 10, Appl
c	23	128.6	4.5	1735	5	PCT-US92-10885-10	Sequence 10, Appl
	24	124.2	4.3	23673	4	US-09-773-816-1	Sequence 1, Appli
c	25	121.6	4.2	1194	4	US-08-765-907A-9	Sequence 9, Appli
c	26	117	4.1	50937	4	US-09-428-517-1	Sequence 1, Appli
c	27	116.6	4.1	1212	6	5212296-17	Patent No. 5212296
	28	115.6	4.0	8438	1	US-07-945-283-1	Sequence 1, Appli
c	29	114.6	4.0	1998	6	5212296-8	Patent No. 5212296
c	30	113.8	4.0	1221	6	5212296-16	Patent No. 5212296
c	31	113.8	4.0	1879	6	5212296-5	Patent No. 5212296
c	32	113.8	4.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	33	109.8	3.8	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	34	108.8	3.8	13842	4	US-09-105-537-30	Sequence 30, Appl
c	35	108.8	3.8	36778	4	US-09-105-537-5	Sequence 5, Appli
c	36	108.8	3.8	38506	3	US-09-320-878-19	Sequence 19, Appl
c	37	108.6	3.8	15079	4	US-09-385-028-1	Sequence 1, Appli
c	38	108.4	3.8	1227	4	US-09-385-028-23	Sequence 23, Appl
c	39	106.8	3.7	15872	4	US-09-105-537-1	Sequence 1, Appli
c	40	104.6	3.6	5392	2	US-08-403-852D-1	Sequence 1, Appli
c	41	104.6	3.6	5392	3	US-08-510-646B-1	Sequence 1, Appli
c	42	104.6	3.6	5392	4	US-09-231-818-1	Sequence 1, Appli
	43	103.4	3.6	2721	6	5215881-2	Patent No. 5215881
c	44	103	3.6	11220	4	US-09-105-537-32	Sequence 32, Appl
	45	102.2	3.6	4257	2	US-08-690-473-1	Sequence 1, Appli

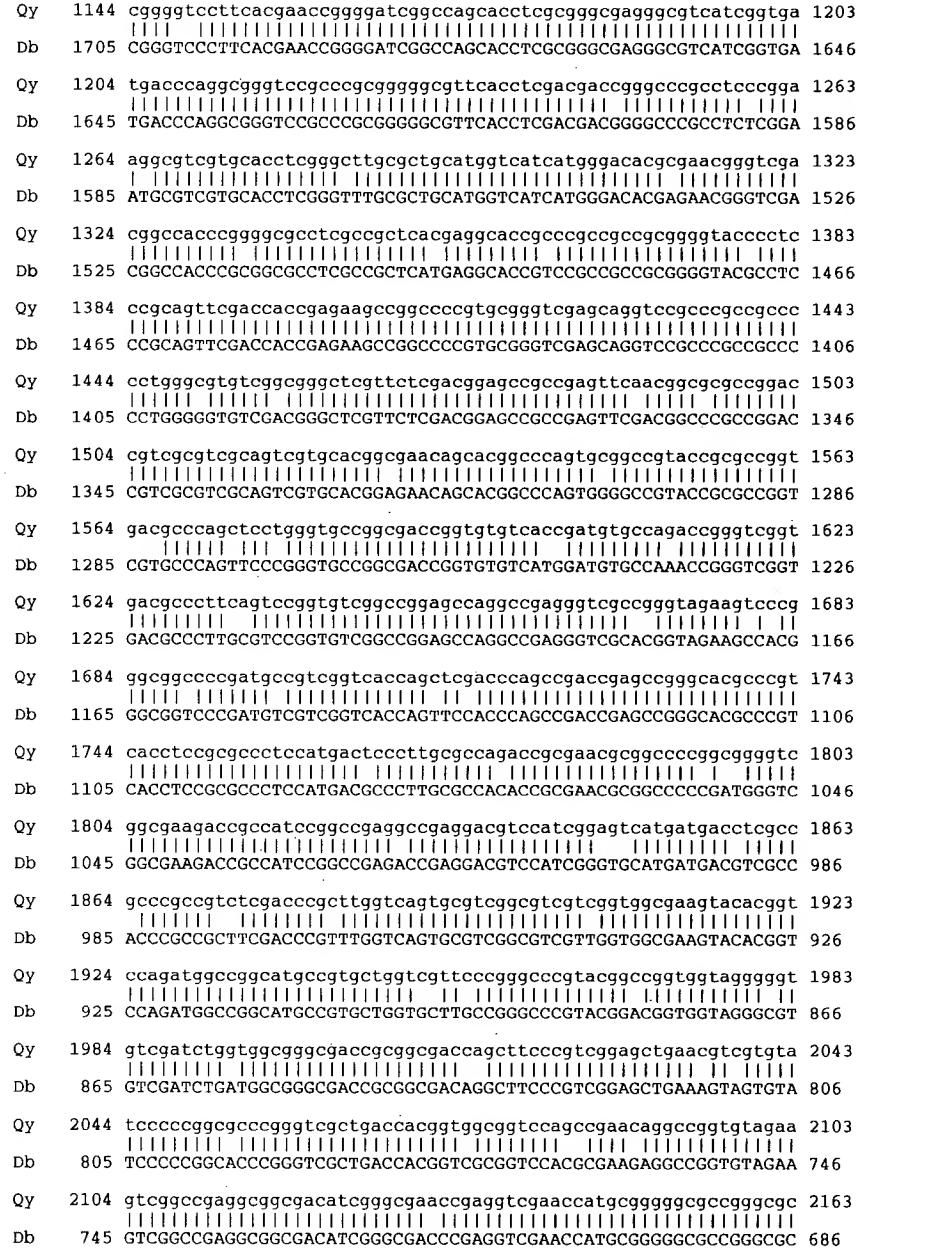
ALIGNMENTS

RESULT 1
US-09-096-982-4/c
; Sequence 4, Application US/09096982
; Patent No. 5962293
; GENERAL INFORMATION:
; APPLICANT: Strohl, William R.
; APPLICANT: Dickens, Michael L.
; APPLICANT: DeSanti, Charles L.
; TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CALFEE, HALTER & GRISWOLD
; STREET: 800 Superior Avenue, Suite 1400
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096,982
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22727/00131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 216-622-8458
; TELEFAX: 216-241-0816
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1498..2764
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1498..2764
US-09-096-982-4

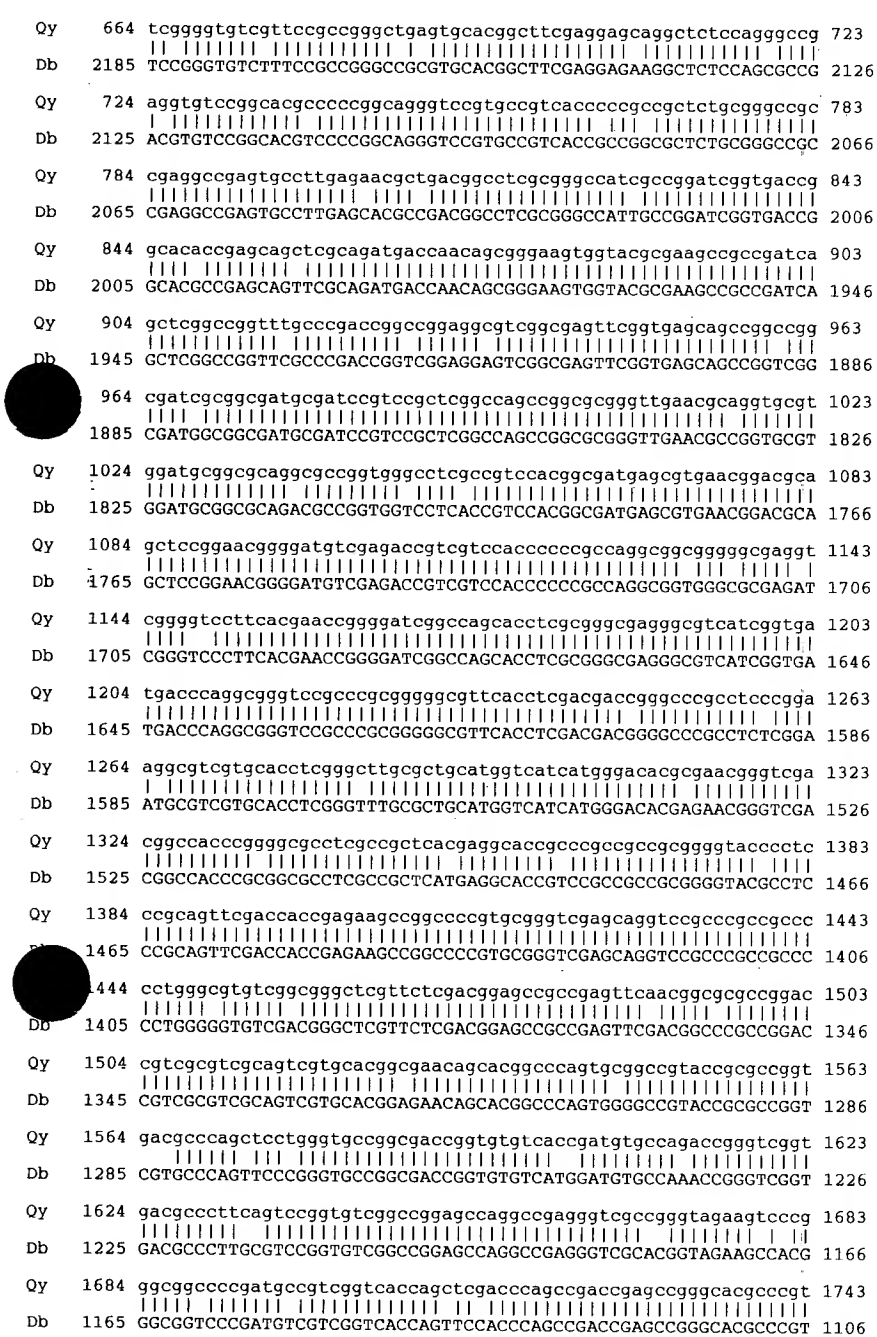
Query Match 88.6%; Score 2542.2; DB 2; Length 3196;

Best Local Similarity 93.6%; Pred. No. 0;
Matches 2681; Conservative 0; Mismatches 163; Indels 21; Gaps 2;





```
Qy   2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632

Qy   2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
                    |||||||||||||||| |||||||||||| || ||||||| |||||
Db    633 -----------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGGCCCGCCGCGGTCAGC 586

Qy   2284 gcgcggtgtcgccgacggcggccgcgccggcctcccagagcttcgccgcgaggccggcgt 2343
          ||||||||| |||||||||||||| |||||||||||||| ||||||||||||||||||
Db    585 GCGCGGTGTGGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526

Qy   2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCA 466

Qy   2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
          |||||||||||||||| |||||||||||||| |||||||||||||| |||||||| |
Db    465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406

Qy   2464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
          |||||||||||||||||||||||||| ||||||||| |||||||| ||||||||||||
Db    405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346

Qy   2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
          |||||||| |||||||| |||||||||| |||||||||||||||||||| ||||||||||
Db    345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286

Qy   2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
            ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226

Qy   2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166

Qy   2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||| |
Db    165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCCG 106

Qy   2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
          |||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46

      2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868
           |||||||||||||||||| |||||||||||||||||||||||||
        45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1


RESULT   2
US-08-653-650A-4/c
; Sequence 4, Application US/08653650A
; Patent No 5976830
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/653,650A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3196 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1498..2764
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1498..2764
US-08-653-650A-4

  Query Match           88.6%;  Score 2542.2;  DB 2;  Length 3196;
  Best Local Similarity 93.6%;  Pred. No. 0;
  Matches 2681;  Conservative    0;  Mismatches  163;  Indels   21;  Gaps    2;

Qy      5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
          |||||||||| |||||||||||||||||||||||| || ||||||||||||||||||
Db   2845 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2786

Qy     65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
          || |||||||| |||| |||| |||||||||||||||| || |||||||||| ||||
Db   2785 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2726

Qy    124 gggccccctt ccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
          |||||||||||| |||||||||||||| |||||||||||||||||||| | |||||
Db   2725 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2666

Qy    184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2665 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2606

Qy    244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
          |||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||||
Db   2605 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2546

Qy    304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2545 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2486

Qy    364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
          |||||||||||||||||||||||| |||| ||||  |||||||||||||||||||||||
Db   2485 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2426

Qy    424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
          |||||| |||||||||||||||||| |||||||||||||||||||||||||||||| ||
Db   2425 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2366

Qy    484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
          ||||| ||||||| |||||||||| |||||||||||||||||||| ||||||||||||
Db   2365 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2306

Qy    544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2305 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2246

Qy    604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
          |||||||||||||||||||||||| ||||| ||||| || || ||||||||||||||||
Db   2245 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2186
```


Qy 664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
Db 2185 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 2126
Qy 724 aggtgtccggcacgccccccggcagggtccgtgccgtcaccccgccgctctgcgggccgc 783
Db 2125 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 2066
Qy 784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
Db 2065 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 2006
Qy 844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
Db 2005 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1946
Qy 904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
Db 1945 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1886
Qy 964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
Db 1885 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1826
Qy 1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
Db 1825 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1766
Qy 1084 gctccggaacggggatgtcgagaccgtcgtccacccccgccaggcggcgggggcgaggt 1143
Db 1765 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1706
Qy 1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
Db 1705 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1646
Qy 1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
Db 1645 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 1586
Qy 1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
Db 1585 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 1526
Qy 1324 cggccacccggggcgcctcgccgctcacgaggcacccgcccgccgcggggtaccctc 1383
Db 1525 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCGTCCGCCGCCGCGGGGTACGCCTC 1466
Qy 1384 ccgcagttcgaccaccgagaagccggccccgtgcgggtcgagcaggtccgcccgccgccc 1443
Db 1465 CCGCAGTTCGACCACCGAGAAGCCGGCCCCGTGCGGGTCGAGCAGGTCCGCCCGCCGCCC 1406
Qy 1444 cctgggcgtgtcggcgggctcgttctcgacggagccgccgagttcaacggcgcgccggac 1503
Db 1405 CCTGGGGGTGTCGACGGGCTCGTTCTCGACGGAGCCGCCGAGTTCGACGGCCCGCCGGAC 1346
Qy 1504 cgtcgcgtcgcagtcgtgcacggcgaacagcacggcccagtgcggccgtaccgcgccggt 1563
Db 1345 CGTCGCGTCGCAGTCGTGCACGGAGAACAGCACGGCCCAGTGGGGCCGTACCGCGCCGGT 1286
Qy 1564 gacgcccagctcctgggtgccggcgaccggtgtgtcaccgatgtgccagaccgggtcggt 1623
Db 1285 CGTGCCCAGTTCCCGGGTGCCGGCGACCGGTGTGTCATGGATGTGCCAAACCGGGTCGGT 1226
Qy 1624 gacgcccttcagtccggtgtcggccggagccaggccgagggtcgccgggtagaagtcccg 1683
Db 1225 GACGCCCTTGCGTCCGGTGTCGGCCGGAGCCAGGCCGAGGGTCGCACGGTAGAAGCCACG 1166
Qy 1684 ggcggccccgatgccgtcggtcaccagctcgacccagccgaccgagccgggcacgcccgt 1743
Db 1165 GGCGGTCCCGATGTCGTCGGTCACCAGTTCCACCCAGCCGACCGAGCCGGGCACGCCCGT 1106
Qy 1744 cacctccgcgccctccatgactcccttgcgccagaccgcgaacgcggccccggcggggtc 1803
Db 1105 CACCTCCGCGCCCTCCATGACGCCCTTGCGCCACACCGCGAACGCGGCCCCCGATGGGTC 1046
Qy 1804 ggcgaagaccgccatccggccgaggccgaggacgtccatcggagtcatgatgacctcgcc 1863
Db 1045 GGCGAAGACCGCCATCCGGCCGAGACCGAGGACGTCCATCGGGTGCATGATGACGTCGCC 986
Qy 1864 gcccgccgtctcgacccgcttggtcagtgcgtcggcgtcgtcggtggcgaagtacacggt 1923
Db 985 ACCCGCCGCTTCGACCCGTTTGGTCAGTGCGTCGGCGTCGTTGGTGGCGAAGTACACGGT 926
Qy 1924 ccagatggccggcatgccgtgctggtcgttcccgggcccgtacggccggtggtaggggt 1983
Db 925 CCAGATGGCCGGCATGCCGTGCTGGTGCTTGCCGGGCCCGTACGGACGGTGGTAGGGCGT 866
Qy 1984 gtcgatctggtggcgggcgaccgcggcgaccagcttcccgtcggagctgaacgtcgtgta 2043
Db 865 GTCGATCTGATGGCGGGCGACCGCGGCGACAGGCTTCCCGTCGGAGCTGAAAGTAGTGTA 806
Qy 2044 tcccccggcgcccgggtcgctgaccacggtggcggtccagccgaacaggccggtgtagaa 2103
Db 805 TCCCCCGGCACCCGGGTCGCTGACCACGGTCGCGGTCCACGCGAAGAGGCCGGTGTAGAA 746
Qy 2104 gtcggccgaggcggcgacatcgggcgaaccgaggtcgaaccatgcgggggcgccgggcgc 2163
Db 745 GTCGGCCGAGGCGGCGACATCGGGCGACCCGAGGTCGAACCATGCGGGGGCGCCGGGCGC 686
Qy 2164 gaacctggtcacgaatcgttcctttcgatggatcggcacacgagcgtctgcgctcgcgga 2223
Db 685 GAACCTGGTCACGAATCGTTCCTTTCGACGGATCGGCACACGAGCGTCTGCG-------- 632
Qy 2224 tgagacggacatctcgcggatgagacggacatgcgggcggggcgggccgccgccgtcagt 2283
Db 633 ----------CTCGCGGATGAGACGGGCATGCGGGCGGGGAGGCCCGCCGCGGTCAGC 586
Qy 2284 gcgcggtgtcgccgacggcggccgccgccggcctcccagagcttcgccgcgaggccggcgt 2343
Db 585 GCGCGGTGTGGCCGACGGCGGCCGCGCTGGCCTCCCAGAGCTTGGCCGCGAGGCCGGCGT 526
Qy 2344 cggcggtcgggccgctcaccggggacagccgccggtcgctgtagtagccgcccgtggtca 2403
Db 525 CGGCGGTCGGGCCGCTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCA 466
Qy 2404 actcctcggccggcgcggacgccagccacacgagggtgtcggcgcccttcgccgcggagc 2463
Db 465 ACTCCTCGGCCGGCGCGGCCGCCAGCCACACGAGAGTGTCGGCGCCCTTGGCCGCGGACC 406
Qy 2464 gcaggaaggggttgaaccggaagtaggacgaggcgaccgtgccccgtccgatgcgggtgc 2523
Db 405 GCAGGAAGGGGTTGAACCGGAAGTAGGTCGAGGCGACTGTGCCCCGCCCGATGCGGGTGC 346
Qy 2524 ggacctcaccggggtgatagctgaccgccagcacgtccggccagcgcctggcggcctccg 2583
Db 345 GGACCTCGCCGGGGTGGTAGCTGACCGTCAGCACGTCCGGCCAGCGCCGGGCGGCCTCCG 286
Qy 2584 ccgcggtcatgatgttggcctgtttggacgtgccgtacgcctggccggcgctgtagcggt 2643
Db 285 TGGCGGTCATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCCGCGCTGTAGCGGT 226
Qy 2644 gacggtcgccgttgaggtcgtccgggtcgatccggccctgggtgtacgcgtcggacgagg 2703
Db 225 GACGGTCGCCGTTGAGATCGTCCGGGTCGATCCGGCCCTGGGTGTACGCGTCGGACGAGG 166
Qy 2704 tgaggatcagccgcccgcccgcgagccgctcccgcagcagccgtgccagcaggaagcctg 2763
Db 165 TGAGGATCAACCGCCCGCCCGCGAGCCGCTCCCGCAGCAGCCGTGCCAGCAGGAAGCCCG 106
Qy 2764 cgaggtgattgacctggatggtggcctcgaacccgtcctgggtcgtggtgcgcgaccaga 2823
Db 105 CGAGATGATTGACCTGGAGGGTGGCCTCGAACCCGTCCTGGGTCGTGGTGCGCGACCAGA 46
Qy 2824 acatgccgccggcgttgctggccatgacatcgatgcgcgggtacc 2868

```
                  |||||||||||||||||| ||||||||||||||||||||||||||
Db      45 ACATGCCGCCGGCGTTGCCGGCCATGACATCGATGCGCGGGTACC 1

RESULT   3
US-08-396-218-1/c
; Sequence 1, Application US/08396218
; Patent No. 5695966
;  GENERAL INFORMATION:
;    APPLICANT:  INVENTI, Augusto
;    APPLICANT:  BREME, Umberto
;    APPLICANT:  COLOMBO, Anna L
;    APPLICANT:  HUTCHINSON, Charles R
;    APPLICANT:  OTTEN, Sharee
;    APPLICANT:  SCOTTI, Claudio
;    TITLE OF INVENTION:  PROCESS FOR PREPARING DOXORUBICIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIKAIDO, MARMELSTEIN, MURRAY & ORAM
;      STREET:  655 Fifteenth Street, N. W., Suite 330 - G
;      STREET:  Street Lobby
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,218
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KITTS, Monica C
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-5002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202/638-5000
;      TELEFAX:  202/638-4810
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1269 base pairs
;      TYPE:  nucleic acid
       STRANDEDNESS:  double
       TOPOLOGY:  linear
     MOLECULE TYPE:  DNA (genomic)
     HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces peucetius
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1269
US-08-396-218-1


  Query Match           44.2%;  Score 1269;  DB 1;  Length 1269;
  Best Local Similarity  100.0%;  Pred. No. 3e-191;
  Matches 1269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 tcagcgcagccagacgggcagttcggtgagccgcgccgtctgggccccccttccggcacca 142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1269 TCAGCGCAGCCAGACGGGCAGTTCGGTGAGCCGCGCCGTCTGGGCCCCCTTCCGGCACCA 1210

Qy     143 ccgcaactcgtcgtacggcacggccagtcgggcctcggggaacctgctgcgcagtacgcc 202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1209 CCGCAACTCGTCGTACGGCACGGCCAGTCGGGCCTCGGGGAACCTGCTGCGCAGTACGCC 1150

Qy     203 gatcatcgtgcgcgactccagctgggcgagctgctccccgatgcagtagtgcggcccgtc 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1149 GATCATCGTGCGCGACTCCAGCTGGGCGAGCTGCTCCCCGATGCAGTAGTGCGGCCCGTC 1090

Qy     263 gccgaaggtgagccgccgccacgagggacggtccgggtggaaggcgtgcggggcgtcgtg 322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1089 GCCGAAGGTGAGCCGCCGCCACGAGGGACGGTCCGGGTGGAAGGCGTGCGGGGCGTCGTG 1030

Qy     323 atggcggccgtcggtgttggtgccctcgatgtccaccagcaccggcgctccgcggggcag 382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1029 ATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCCGCGGGGCAG 970

Qy     383 ccggacgccgccgatggtcacctccgtggcagcgaacctccacaacgtgtagggcaccgg 442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     969 CCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACCTCCACAACGTGTAGGGCACCGG 910

Qy     443 cgggtggtagcgcagcgcctcctccacgaaccgggagacggcgtcctcgtcggcatccgc 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     909 CGGGTGGTAGCGCAGCGCCTCCTCCACGAACCGGGAGACGGCGTCCTCGTCGGCATCCGC 850

Qy     503 cgcgaggcggcccgccaggacctccgcgagcaggaagcccaggaaggagccggtggtgtc 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     849 CGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGGAGCCGGTGGTGTC 790

Qy     563 gtggccggcgaagatgagcccggtgatcatgtagacgagctggtcgtcggagaccgagcc 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     789 GTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACGAGCTGGTCGTCGGAGACCGAGCC 730

Qy     623 gaactcggcctgcgcgcgctcgtacagcacgcgggtcatggtcggggtgtcgttccgccg 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     729 GAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCATGGTCGGGGTGTCGTTCCGCCG 670

Qy     683 ggctgagtgcacggcttcgaggagcaggctctccagggccgaggtgtccggcacgccccc 742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 GGCTGAGTGCACGGCTTCGAGGAGCAGGCTCTCCAGGGCCGAGGTGTCCGGCACGCCCCC 610

Qy     743 ggcagggtccgtgccgtcaccccgccgctctgcgggccgccgaggccgagtgccttgag 802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 GGCAGGGTCCGTGCCGTCACCCCCGCCGCTCTGCGGGCCGCCGAGGCCGAGTGCCTTGAG 550

Qy     803 aacgctgacggcctcgcgggccatcgccggatcggtgaccggcacaccgagcagctcgca 862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 AACGCTGACGGCCTCGCGGGCCATCGCCGGATCGGTGACCGGCACACCGAGCAGCTCGCA 490

Qy     863 gatgaccaacagcgggaagtggtacgcgaagccgccgatcagctcggccggtttgcccga 922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 GATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCAGCTCGGCCGGTTTGCCCGA 430

Qy     923 ccggccggaggcgtcggcgagttcggtgagcagccggccggcgatcgcggcgatgcgatc 982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 CCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCCGGCGATCGCGGCGATGCGATC 370

Qy     983 cgtccgctcggccagccggcgcgggttgaacgcaggtgcgtggatgcggcgcaggcgccg 1042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 CGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTGCGTGGATGCGGCGCAGGCGCCG 310

Qy    1043 gtgggcctcgccgtccacggcgatgagcgtgaacggacgcagctccggaacggggatgtc 1102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 GTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGACGCAGCTCCGGAACGGGGATGTC 250

Qy    1103 gagaccgtcgtccacccccccgccaggcggcgggggcgaggtcggggtccttcacgaaccg 1162
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 GAGACCGTCGTCCACCCCCCGCCAGGCGGCGGGGGCGAGGTCGGGGTCCTTCACGAACCG 190

Qy    1163 gggatcggccagcacctcgcgggcgagggcgtcatcggtgatgacccaggcgggtccgcc 1222
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 GGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGATGACCCAGGCGGGTCCGCC 130

Qy    1223 cgcgggggcgttcacctcgacgaccgggcccgcctcccggaaggcgtcgtgcacctcggg 1282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 CGCGGGGGCGTTCACCTCGACGACCGGGCCCGCCTCCCGGAAGGCGTCGTGCACCTCGGG 70

Qy    1283 cttgcgctgcatggtcatcatgggacacgcgaacgggtcgacggccacccggggcgcctc 1342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 CTTGCGCTGCATGGTCATCATGGGACACGCGAACGGGTCGACGGCCACCCGGGGCGCCTC 10
```

```
Qy   1343 gccgctcac 1351
          |||||||||
Db      9 GCCGCTCAC 1

RESULT   4
US-08-760-116-1/c
; Sequence 1, Application US/08760116
; Patent No. 5786190
; GENERAL INFORMATION:
;    APPLICANT:  INVENTI, Augusto
;    APPLICANT:  BREME, Umberto
;    APPLICANT:  COLOMBO, Anna L
;    APPLICANT:  HUTCHINSON, Charles R
;    APPLICANT:  OTTEN, Sharee
;    APPLICANT:  SCOTTI, Claudio
;    TITLE OF INVENTION:  PROCESS FOR PREPARING DOXORUBICIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIKAIDO, MARMELSTEIN, MURRAY & ORAM LLP
;      STREET:  655 Fifteenth Street, N. W., Suite 330 - G
;      STREET:  Street Lobby
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/760,116
;      FILING DATE:  3-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,218
;      FILING DATE:  27-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KITTS, Monica C
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-6007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202/638-5000
;      TELEFAX:  202/638-4810
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1269 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
     MOLECULE TYPE:  DNA (genomic)
     HYPOTHETICAL:  NO
     ANTI-SENSE:  NO
     ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces peucetius
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1269
US-08-760-116-1


  Query Match             44.2%;  Score 1269;  DB 1;  Length 1269;
  Best Local Similarity   100.0%;  Pred. No. 3e-191;
  Matches 1269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 tcagcgcagccagacgggcagttcggtgagccgcgccgtctgggccccttccggcacca 142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1269 TCAGCGCAGCCAGACGGGCAGTTCGGTGAGCCGCGCCGTCTGGGCCCCTTCCGGCACCA 1210

Qy     143 ccgcaactcgtcgtacggcacggccagtcgggcctcggggaacctgctgcgcagtacgcc 202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1209 CCGCAACTCGTCGTACGGCACGGCCAGTCGGGCCTCGGGGAACCTGCTGCGCAGTACGCC 1150

Qy     203 gatcatcgtgcgcgactccagctgggcgagctgctccccgatgcagtagtgcggcccgtc 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1149 GATCATCGTGCGCGACTCCAGCTGGGCGAGCTGCTCCCCGATGCAGTAGTGCGGCCCGTC 1090

Qy     263 gccgaaggtgagccgccgccacgagggacggtccgggtggaaggcgtgcggggcgtcgtg 322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1089 GCCGAAGGTGAGCCGCCGCCACGAGGGACGGTCCGGGTGGAAGGCGTGCGGGGCGTCGTG 1030

Qy     323 atggcggccgtcggtgttggtgccctcgatgtccaccagcaccggcgctccgcgggcag 382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1029 ATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCCGCGGGGCAG 970

Qy     383 ccggacgccgccgatggtcacctccgtggcagcgaacctccacaacgtgtagggcaccgg 442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     969 CCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACCTCCACAACGTGTAGGGCACCGG 910

Qy     443 cgggtggtagcgcagcgcctcctccacgaaccgggagacggcgtcctcgtcggcatccgc 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     909 CGGGTGGTAGCGCAGCGCCTCCTCCACGAACCGGGAGACGGCGTCCTCGTCGGCATCCGC 850

Qy     503 cgcgaggcggcccgccaggacctccgcgagcaggaagcccaggaaggagccggtggtgtc 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     849 CGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGGAGCCGGTGGTGTC 790

Qy     563 gtggccggcgaagatgagcccggtgatcatgtagacgagctggtcgtcggagaccgagcc 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     789 GTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACGAGCTGGTCGTCGGAGACCGAGCC 730

Qy     623 gaactcggcctgcgcgcgctcgtacagcacgcgggtcatggtcggggtgtcgttccgccg 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     729 GAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCATGGTCGGGGTGTCGTTCCGCCG 670

Qy     683 ggctgagtgcacggcttcgaggagcaggctctccagggccgaggtgtccggcacgccccc 742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 GGCTGAGTGCACGGCTTCGAGGAGCAGGCTCTCCAGGGCCGAGGTGTCCGGCACGCCCCC 610

Qy     743 ggcagggtccgtgccgtcaccccgccgctctgcgggccgccgaggccgagtgccttgag 802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 GGCAGGGTCCGTGCCGTCACCCCCGCCGCTCTGCGGGCCGCCGAGGCCGAGTGCCTTGAG 550

Qy     803 aacgctgacggcctcgcgggccatcgccggatcggtgaccggcacaccgagcagctcgca 862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 AACGCTGACGGCCTCGCGGGCCATCGCCGGATCGGTGACCGGCACACCGAGCAGCTCGCA 490

Qy     863 gatgaccaacagcgggaagtggtacgcgaagccgccgatcagctcggccggtttgcccga 922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 GATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCAGCTCGGCCGGTTTGCCCGA 430

Qy     923 ccggccggaggcgtcggcgagttcggtgagcagccggccggcgatcgcggcgatgcgatc 982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 CCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCCGGCGATCGCGGCGATGCGATC 370

Qy     983 cgtccgctcggccagccggcgcgggttgaacgcaggtgcgtggatgcggcgcaggcgccg 1042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 CGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTGCGTGGATGCGGCGCAGGCGCCG 310

Qy    1043 gtgggcctcgccgtccacggcgatgagcgtgaacggacgcagctccggaacggggatgtc 1102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 GTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGACGCAGCTCCGGAACGGGGATGTC 250

Qy    1103 gagaccgtcgtccacccccccgccaggcggcgggggcgaggtcgggggtccttcacgaaccg 1162
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 GAGACCGTCGTCCACCCCCCCGCCAGGCGGCGGGGGCGAGGTCGGGGTCCTTCACGAACCG 190

Qy    1163 gggatcggccagcacctcgcgggcgagggcgtcatcggtgatgacccaggcgggtccgcc 1222
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 GGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGATGACCCAGGCGGGTCCGCC 130

Qy    1223 cgcgggggcgttcacctcgacgaccgggcccgcctcccggaaggcgtcgtgcacctcggg 1282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 CGCGGGGGCGTTCACCTCGACGACCGGGCCCGCCTCCCGGAAGGCGTCGTGCACCTCGGG 70
```

```
Qy   1283 cttgcgctgcatggtcatcatgggacacgcgaacgggtcgacggccacccggggcgcctc 1342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 CTTGCGCTGCATGGTCATCATGGGACACGCGAACGGGTCGACGGCCACCCGGGGCGCCTC 10

Qy   1343 gccgctcac 1351
          |||||||||
Db      9 GCCGCTCAC 1


RESULT   5
US-09-096-982-7/c
; Sequence 7, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
     APPLICANT:  Strohl, William R.
     APPLICANT:  Dickens, Michael L.
     APPLICANT:  DeSanti, Charles L.
     TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/096,982
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2081 base pairs
       TYPE:  nucleic acid
       STRANDEDNESS:  double
       TOPOLOGY:  linear
     MOLECULE TYPE:  DNA (genomic)
     FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  227..1649
US-09-096-982-7


  Query Match             42.6%;  Score 1221.6;  DB 2;  Length 2081;
  Best Local Similarity   94.1%;  Pred. No. 7.7e-184;
  Matches 1280;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy      5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
          ||||||||||    |||||||||||||||||||||||| || ||||||||||||||||||
Db   1730 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 1671

Qy     65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
          || |||||||| ||||  |||| |||||||||||||||| || |||||||||| ||||
Db   1670 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 1611

Qy    124 gggccccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
          |||||||||||| |||||||||||||| |||||||||||||||||||||| | |||||
Db   1610 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 1551

Qy    184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
```

```
Db   1550 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 1491

Qy    244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
          |||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||||
Db   1490 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 1431

Qy    304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   1430 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 1371

Qy    364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
          |||||||||||||||||||||||| |||| ||||  ||||||||||||||||||||||||
Db   1370 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 1311

Qy    424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
          |||||| |||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db   1310 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 1251

Qy    484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
          ||||| ||||||| |||||||||||| |||||||||||||||||||| ||||||||||||
Db   1250 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 1191

Qy    544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
          |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1190 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 1131

Qy    604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
          |||||||||||||||||||||||| ||||| ||||| || || ||||||||||||||||
Db   1130 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 1071

Qy    664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
          || |||||| |||||||||| | |||||||||||||||||| |||||||||| ||||
Db   1070 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 1011

Qy    724 aggtgtccggcacgccccccggcagggtccgtgccgtcacccccgccgcctctgcgggccgc 783
          | |||||||||||| |||||||||||||||||||||||||| ||| ||||||||||||||
Db   1010 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 951

Qy    784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
          |||||||||||||||||| |||| |||||||||||||||| ||||||||||||||||||
Db   950 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 891

Qy    844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
          |||| |||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   890 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 831

Qy    904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
          |||||||||||| |||||||||| |||||| |||||||||||||||||||||||||| |||
Db   830 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 771

Qy    964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   770 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 711

Qy   1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
          |||||||||||| |||||||||||| |||| ||||||||||||||||||||||||||||
Db   710 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 651

Qy   1084 gctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcgaggt 1143
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||||.|
Db   650 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 591

Qy   1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
          ||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   590 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 531

Qy   1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   530 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 471

Qy   1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
          | |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db   470 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 411
```

```
Qy   1324 cggccacccggggcgcctcgccgctcacgaggcaccgccc 1363
          ||||||||| ||||||||||||||| |||||||| | |
Db    410 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCCCTC 371


RESULT   6
US-08-653-650A-7/c
; Sequence 7, Application US/08653650A
; Patent No. 5976830
;  GENERAL INFORMATION:
;     APPLICANT:  Strohl, William R.
;     APPLICANT:  Dickens, Michael L.
;     APPLICANT:  DeSanti, Charles L.
;     TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;       STREET:  800 Superior Avenue, Suite 1400
;       CITY:  Cleveland
;       STATE:  Ohio
;       COUNTRY:  USA
;       ZIP:  44114-2688
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/653,650A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Golrick, Mary E.
;       REGISTRATION NUMBER:  34829
;       REFERENCE/DOCKET NUMBER:  22727/00131
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  216-622-8458
;       TELEFAX:  216-241-0816
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2081 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  227..1649
US-08-653-650A-7

ery Match          42.6%;  Score 1221.6;  DB 2;  Length 2081;
t Local Similarity  94.1%;  Pred. No. 7.7e-184;
tches 1280;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy       5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
           |||||||||||    |||||||||||||||||||||||| || ||||||||||||||||||
Db    1730 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 1671

Qy      65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
           || |||||||| ||||  |||| |||||||||||||||| || ||||||||||| ||||
Db    1670 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 1611

Qy     124 gggccccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
           |||||||||||| ||||||||||||| |||||||||||||||||||||| | |||||
Db    1610 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 1551

Qy     184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    1550 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 1491

Qy     244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
           |||||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||||
Db    1490 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 1431

Qy     304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1430 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 1371

Qy     364 ccggcgctccgcggggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
           |||||||||||||||||||||||| |||| ||||  ||||||||||||||||||||||||
Db    1370 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 1311

Qy     424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
           |||||| |||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db    1310 ACAACGAGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 1251

Qy     484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
           ||||| ||||||| |||||||||||| |||||||||||||||||||||||| ||||||||||||
Db    1250 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 1191

Qy     544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
           |||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    1190 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 1131

Qy     604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
           |||||||||||||||||||||||| |||| |||||  || || ||||||||||||||||||
Db    1130 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 1071

Qy     664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
           || ||||||| |||||||||| | |||||||||||||||| |||||||||| ||||
Db    1070 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 1011

Qy     724 aggtgtccggcacgccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
           | |||||||||||| |||||||||||||||||||||||||| ||| ||||||||||||||||
Db    1010 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 951

Qy     784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
           |||||||||||||||||| |||| |||||||||||||||| ||||||||||||||||||||
Db    950 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 891

Qy     844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
           |||| |||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    890 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 831

Qy     904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
           |||||||||||| |||||||||| |||||| |||||||||||||||||||||||||| |||
Db    830 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 771

Qy     964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    770 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 711

Qy    1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
           |||||||||||| ||||||||| |||| ||||||||||||||||||||||||||||||||
Db    710 GGATGCGGCGCAGACGCCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 651

Qy    1084 gctccggaacggggatgtcgagaccgtcgtccacccccccgccaggcggcgggggcgaggt 1143
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |
Db    650 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 591

Qy    1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
           ||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 CGGGTCCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 531

Qy    1204 tgacccaggcgggtccgcccgcgggggcgttcacctcgacgaccgggcccgcctcccgga 1263
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||
Db    530 TGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGACGACGGGGCCCGCCTCTCGGA 471

Qy    1264 aggcgtcgtgcacctcgggcttgcgctgcatggtcatcatgggacacgcgaacgggtcga 1323
           | |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db    470 ATGCGTCGTGCACCTCGGGTTTGCGCTGCATGGTCATCATGGGACACGAGAACGGGTCGA 411

Qy    1324 cggccacccggggcgcctcgccgctcacgaggcaccgccc 1363
           ||||||||| ||||||||||||||| |||||||| | |
```

```
Db    410 CGGCCACCCGCGGCGCCTCGCCGCTCATGAGGCACCCCTC 371


RESULT   7
US-09-096-982-6/c
; Sequence 6, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
       ADDRESSEE:  CALFEE, HALTER & GRISWOLD
       STREET:  800 Superior Avenue, Suite 1400
       CITY:  Cleveland
       STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/09/096,982
       FILING DATE:
       CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
       NAME:  Golrick, Mary E.
       REGISTRATION NUMBER:  34829
       REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
       TELEPHONE:  216-622-8458
       TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
       LENGTH:  3013 base pairs
       TYPE:  nucleic acid
       STRANDEDNESS:  double
       TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-09-096-982-6


     ry Match            38.4%;  Score 1102.2;  DB 2;  Length 3013;
     t Local Similarity  94.4%;  Pred. No. 3.8e-165;
     ches 1154;  Conservative    0;  Mismatches   68;  Indels    1;  Gaps    1;

Qy      5 ccgcaccgggtacacggcacgggaccgcccaccgcgcggtgcgcggtgggcggtcccgtg 64
          ||||||||||    |||||||||||||||||||||||| || ||||||||||||||||||
Db   2662 CCGCACCGGGCGACCGGCACGGGACCGCCCACCGCGCGATGGGCGGTGGGCGGTCCCGTG 2603

Qy     65 cc-ggtcgcggccggcggatcagcgcagccagacgggcagttcggtgagccgcgccgtct 123
          || |||||||| ||||  |||| ||||||||||||||| || |||||||||| ||||
Db   2602 CCGGGTCGCGGTCGGCCCATCAACGCAGCCAGACGGGCAGGTCAGTGAGCCGCGCTGTCT 2543

Qy    124 gggccccccttccggcaccaccgcaactcgtcgtacggcacggccagtcgggcctcgggga 183
          |||||||||||| |||||||||||||| |||||||||||||||||||||| | |||||
Db   2542 GGGCCCCCTTCCTGCACCACCGCAACTCCTCGTACGGCACGGCCAGTCGGGCTTGGGGGA 2483

Qy    184 acctgctgcgcagtacgccgatcatcgtgcgcgactccagctgggcgagctgctccccga 243
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   2482 ACCTGCTGCGCAGTACGCCGATCATCGTGCGCGATTCCAGCTGGGCGAGCTGCTCCCCGA 2423

Qy    244 tgcagtagtgcggcccgtcgccgaaggtgagccgccgccacgagggacggtccgggtgga 303
          |||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||||
Db   2422 TGCAGTAGTGCGGCCCGTCGCCGAAGGTGAGCCGCCGCCTCGAAGGGCGGTCCGGGTGGA 2363

Qy    304 aggcgtgcggggcgtcgtgatggcggccgtcggtgttggtgccctcgatgtccaccagca 363
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2362 AAGCGTGCGGGGCGTCGTGATGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGTA 2303

Qy    364 ccggcgctccgcgggcagccggacgccgccgatggtcacctccgtggcagcgaacctcc 423
          |||||||||||||||||||||||||| |||| ||||  ||||||||||||||||||||||
Db   2302 CCGGCGCTCCGCGGGGCAGCCGGACACCGCGGATGACCACCTCCGTGGCAGCGAACCTCC 2243

Qy    424 acaacgtgtagggcaccggcgggtggtagcgcagcgcctcctccacgaaccgggagacgg 483
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db   2242 ACAACGTGTAGGGCACCGGCGGGTGGTGGCGCAGCGCCTCCTCCACGAACCGGGAGATGG 2183

Qy    484 cgtcctcgtcggcatccgccgcgaggcggcccgccaggacctccgcgagcaggaagccca 543
          ||||| ||||||| |||||||||| |||||||||||||||||||| ||||||||||||
Db   2182 CGTCCCCGTCGGCGTCCGCCGCGAGACGGCCCGCCAGGACCTCCGCAAGCAGGAAGCCCA 2123

Qy    544 ggaaggagccggtggtgtcgtggccggcgaagatgagcccggtgatcatgtagacgagct 603
          |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   2122 GGAACGAGCCGGTGGTGTCGTGGCCGGCGAAGATGAGTCCGGTGATCATGTAGACGAGCT 2063

Qy    604 ggtcgtcggagaccgagccgaactcggcctgcgcgcgctcgtacagcacgcgggtcatgg 663
          |||||||||||||||||||||||| ||||| ||||| || || ||||||||||||||||
Db   2062 GGTCGTCGGAGACCGAGCCGAACTCTGCCTGTGCGCGTTCATAGAGCACGCGGGTCATGG 2003

Qy    664 tcggggtgtcgttccgccgggctgagtgcacggcttcgaggagcaggctctccagggccg 723
          || ||||||| |||||||||| | |||||||||||||||||| |||||||||| ||||
Db   2002 TCCGGGTGTCTTTCCGCCGGGCCGCGTGCACGGCTTCGAGGAGAAGGCTCTCCAGCGCCG 1943

Qy    724 aggtgtccggcacgcccccggcagggtccgtgccgtcacccccgccgctctgcgggccgc 783
          | |||||||||||| |||||||||||||||||||||||||| ||| ||||||||||||||
Db   1942 ACGTGTCCGGCACGTCCCCGGCAGGGTCCGTGCCGTCACCGCCGGCGCTCTGCGGGCCGC 1883

Qy    784 cgaggccgagtgccttgagaacgctgacggcctcgcgggccatcgccggatcggtgaccg 843
          |||||||||||||||||| |||| |||||||||||||||||| ||||||||||||||||
Db   1882 CGAGGCCGAGTGCCTTGAGCACGCCGACGGCCTCGCGGGCCATTGCCGGATCGGTGACCG 1823

Qy    844 gcacaccgagcagctcgcagatgaccaacagcgggaagtggtacgcgaagccgccgatca 903
          |||| ||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1822 GCACGCCGAGCAGTTCGCAGATGACCAACAGCGGGAAGTGGTACGCGAAGCCGCCGATCA 1763

Qy    904 gctcggccggtttgcccgaccggccggaggcgtcggcgagttcggtgagcagccggccgg 963
          |||||||||||| |||||||||| |||||| |||||||||||||||||||||||||||||
Db   1762 GCTCGGCCGGTTCGCCCGACCGGTCGGAGGAGTCGGCGAGTTCGGTGAGCAGCCGGTCGG 1703

Qy    964 cgatcgcggcgatgcgatccgtccgctcggccagccggcgcgggttgaacgcaggtgcgt 1023
          |||| |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   1702 CGATGGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCCGGTGCGT 1643

Qy    1024 ggatgcggcgcaggcgccggtgggcctcgccgtccacggcgatgagcgtgaacggacgca 1083
           |||||||||||| || |||||| |||| ||||||||||||||||||||||||||||||||
Db   1642 GGATGCGGCGCAGACGGCGGTGGTCCTCACCGTCCACGGCGATGAGCGTGAACGGACGCA 1583

Qy    1084 gctccggaacggggatgtcgagaccgtcgtccacccccgccaggcggcgggggcgaggt 1143
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||| |
Db   1582 GCTCCGGAACGGGGATGTCGAGACCGTCGTCCACCCCCCGCCAGGCGGTGGGCGCGAGAT 1523

Qy    1144 cggggtccttcacgaaccggggatcggccagcacctcgcgggcgagggcgtcatcggtga 1203
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1522 CGGGGTCCTTCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGA 1463

Qy    1204 tgacccaggcgggtccgcccgcg 1226
           |||||||||||||||||||||||
Db   1462 TGACCCAGGCGGGTCCGCCCGCG 1440


Search completed: June 11, 2002, 17:59:11
Job time: 3998 sec
```

OM of: US-09-673-254-1 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 11, 2002 8:07 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09673254/runat_11062002_114210_866/app_query.fasta_1.2973
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=n2p.rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
[redacted]ANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
_MAX=100 -THR_MIN=0 -ALIGN=7 -MODE=LOCAL -OUTFMT=pfs
M=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
SER=US09673254_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-673-254-1
Query length: 2870
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 102.040000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation	..		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAW00729	-	2209.00	2481.39	1.0e-130	42				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36132	-	2093.50	2350.17	1.9e-123	47				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36128	-	2076.00	2331.56	2.3e-122	42				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36134	-	1385.00	1557.20	4.8e-79	27				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW68559	-	978.00	1098.30	1.7e-53	28				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36133	-	961.00	1082.93	1.8e-52	19				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW68560	-	593.00	673.13	2.0e-29	11				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR93254	+	491.00	537.18	8.1e-23	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW57652	+	491.00	537.18	8.1e-23	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR37741	+	491.00	537.17	8.1e-23	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR93254	-	484.00	529.30	2.2e-22	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW57652	-	484.00	529.30	2.2e-22	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR37741	-	484.00	529.29	2.2e-22	10				
[redacted]S1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB71434	-	475.50	512.44	9.0e-22	22				
[redacted]1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR89472	-	464.50	506.95	3.8e-21	11				
[redacted]1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR89469	-	464.50	506.43	3.8e-21	11				
[redacted]S1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84537	-	464.50	506.43	3.8e-21	11				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR89470	-	464.50	506.42	3.8e-21	11				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR89471	-	464.50	504.80	4.0e-21	13				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84541	-	462.50	505.14	5.0e-21	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503	-	461.50	502.71	5.9e-21	12				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY56800	-	461.00	500.70	6.6e-21	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	-	460.00	498.28	7.9e-21	16				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84544	-	459.50	501.76	7.7e-21	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84540	-	459.50	501.32	7.8e-21	11				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84538	-	459.50	500.78	7.9e-21	11				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84403	-	457.00	498.93	1.1e-20	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW68485	-	457.00	495.84	1.2e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY84539	-	456.50	495.78	1.3e-20	13				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB82454	-	456.00	494.71	1.4e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE02533	-	456.00	494.70	1.4e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE02534	-	456.00	494.70	1.4e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE02532	-	454.50	493.03	1.7e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504	-	454.00	495.53	1.7e-20	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU14136	-	453.00	491.33	2.1e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365	+	452.50	493.51	2.1e-20	10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE02535	-	452.00	490.30	2.4e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE02537	-	452.00	490.19	2.4e-20	14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM24516	-	449.50	475.35	4.6e-20	51				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827	-	449.00	486.03	3.8e-20	15				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03533	-	448.00	490.34	3.9e-20					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR28916	-	447.00	486.50	4.8e-20					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB26239	-	446.50	485.65	5.2e-20					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	-	444.50	485.23	6.7e-20					
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22679	-	442.50	479.16	9.7e-20					

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAW00729

seq_documentation_block:
ID AAW00729 standard; Protein; 422 AA.
XX
AC AAW00729;
XX
DT 03-JAN-1997 (first entry)
XX
DE Daunorubicin 14-hydroxylase.
XX
KW dxrA; doxorubicin; bioconversion; anthracycline; antitumour therapy;
KW daunorubicin 14-hydroxylase.
XX
OS Streptomyces peucetius strain 29050.
XX
PN WO9627014-A1.
XX
PD 06-SEP-1996.
XX
PF 20-FEB-1996; 96WO-EP00692.
XX
PR 27-FEB-1995; 95US-0396218.
XX
PA (PHAA) PHARMACIA SPA.
XX
PI Breme U, Colombo AL, Hutchinson CR, Otten S, Scotti C;
PI Solari INVENTI A, Inventi AS;
XX
DR WPI; 1996-412780/41.
DR N-PSDB; AAT33617.
XX
PT New DNA encoding daunorubicin 14-hydroxylase - useful for converting
PT daunorubicin to the antitumour agent doxorubicin
XX
PS Claim 4; Page 15-18; 31pp; English.
XX
CC The present sequence is that of the daunorubicin 14-hydroxylase,
CC isolated from Streptomyces peucetius strain 29050. The enzyme converts
CC daunorubicin to doxorubicin, an anthracycline used widely as an
CC antitumour agent. A vector contg. the drxA gene (AAT33617) could be used
CC to transform a suitable host cell which produces daunorubicin, in order
CC to enhance bioconversion of daunorubicin to doxorubicin. Bioconversion
CC can be carried out either by using directly the free or immobilised
CC transformed cells or by isolating the enzyme, which can be used in the
CC free form or immobilised to resins, glass, cellulose or similar
CC substances by ionic or covalent bonds, or grafted to fibres permeable to
CC the substrate or insolubilised by cross-linkage.
XX
SQ Sequence 422 AA;

alignment_scores:
Quality: 2209.00 Length: 422
Ratio: 5.235 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-673-254-1/rev x AAW00729 ..

Align seg 1/1 to: AAW00729 from: 1 to: 422

1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
|||||||||||||||||||||||||||||||||||||||||||||||||||
1 ValSerGlyGluAlaProArgValAlaValAspProPheAlaCysProMe 17

1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

 1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50

 1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValLy 67

 1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGGTGGACGACGGTCTCG 1102
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   67 sAspProAspLeuAlaProAlaAlaTrpArgGlyValAspAspGlyLeuA 84

 1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

 1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  101 GluAlaHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117

 1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134

  951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  134 euThrGluLeuAlaAspAlaSerGlyArgSerGlyLysProAlaGluLeu 150

  901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  151 IleGlyGlyPheAlaTyrHisPheProLeuLeuValIleCysGluLeuLe 167

  851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValSerValL 184

  801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  184 euLysAlaLeuGlyLeuGlyGlyProGlnSerGlyGlyGlyAspGlyThr 200

  751 GACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  201 AspProAlaGlyGlyValProAspThrSerAlaLeuGluSerLeuLeuLe 217

  701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  217 uGluAlaValHisSerAlaArgArgAsnAspThrProThrMetThrArgV 234

  651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250

  601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  251 LeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrGl 267

  551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284

  501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  284 laAspAlaAspGluAspAlaValSerArgPheValGluGluAlaLeuArg 300

  451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  301 TyrHisProProValProTyrThrLeuTrpArgPheAlaAlaThrGluVa 317

  401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  317 lThrIleGlyGlyValArgLeuProArgGlyAlaProValLeuValAspI 334

  351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  334 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350

  301 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  351 HisProAspArgProSerTrpArgArgLeuThrPheGlyAspGlyProHi 367

  251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384

  201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  384 lyValLeuArgSerArgPheProGluAlaArgLeuAlaValProTyrAsp 400

  151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  401 GluLeuArgTrpCysArgLysGlyAlaGlnThrAlaArgLeuThrGluLe 417

  101 GCCCGTCTGGCTGCGC 86
      ||||||||||||||||
  417 uProValTrpLeuArg 422
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36132

```
seq_documentation_block:
ID   AAW36132 standard; Protein; 474 AA.
XX
AC   AAW36132;
XX
DT   21-MAY-1998  (first entry)
XX
DE   N-terminal modified daunomycin C-14 hydroxylase.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy; plasmid pANT199.
XX
OS   Chimeric - Streptomyces sp. strain C5.
OS   Chimeric - synthetic.
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   31
FT                   /note= "enterokinase cleavage site"
FT   Protein         53..474
FT                   /note= "native daunomycin C-14 hydroxylase"
XX
PN   WO9744439-A2.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08690.
XX
PR   24-MAY-1996;   96US-0653650.
XX
PA   (OHIS ) UNIV OHIO STATE. RES FOUND.
XX
PI   DeSanti CL,  Dickens ML,  Strohl WA;
XX
DR   WPI; 1998-018495/02.
DR   N-PSDB; AAV01452.
XX
PT   Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT   - also hydroxylation and oxidation of other anthracycline(s) with
PT   the same enzyme
XX
PS   Disclosure; Fig 9; 59pp; English.
XX
CC   This protein comprises an N-terminal modified enzyme derived from
CC   the daunomycin C-14 hydroxylase (see AAW36128) of Streptomyces sp.
```

```
CC   strain C5.  It is encoded by a gene construct (see AAV01452) in
CC   plasmid pANT199.  In this plasmid, the doxA gene (see AAV01447) is
CC   translationally fused with a leader sequence encoding 6
CC   histidine residues so that the fusion protein can be affinity
CC   purified on a nickel-agarose gel.  Daunomycin C-14 hydroxylase
CC   is a P450-like enzyme capable of converting daunomycin to the
CC   anticancer agent doxorubicin.  Host cells, especially Streptomyces
CC   host cells, transformed with pANT199 can be used in methods for
CC   the production of doxorubicin from daunomycin or for the
CC   hydroxylation and oxidation of other anthracyclines.
XX
SQ   Sequence   474 AA;

alignment_scores:
          Quality: 2093.50              Length:    432
            Ratio:  4.973                Gaps:      1
[redacted]ent Similarity:  97.454   Percent Identity: 93.056

alignment_block:
 US-09-673-254-1/rev x AAW36132   ..

 Align seg 1/1  to: AAW36132  from: 1  to: 474

    1378 GTACCCCGCGGCGGCGGGCGG...TGCCTCGTGAGCGGCGAGGCGCCCCG 1332
         ||||||   |||:::::::::   ||||||:::||||||||||||||||
      43 ValProTyrGlyAsnSerGluGlyCysLeuMetSerGlyGluAlaProAr 59

    1331 GGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGC 1282
         ||||||||||||||||||:::||||||||||||||||||||||||||||
      59 gValAlaValAspProPheSerCysProMetMetThrMetGlnArgLysP 76

    1281 CCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAAC 1232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      76 roGluValHisAspAlaPheArgGluAlaGlyProValValGluValAsn 92

    1231 GCCCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCG 1182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      93 AlaProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaAr 109

    1181 CGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCG 1132
         ||||||||||||||||||||||||||||||   ||||||||||||||:
     109 gGluValLeuAlaAspProArgPheValLysGlyProAspLeuAlaProT 126

    1131 CCGCCTGGCGGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTG 1082
         ::||||||||||||||||||||||||||||||||||||||||||||||||
     126 hrAlaTrpArgGlyValAspAspGlyLeuAspIleProValProGluLeu 142

    1081 CGTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCG 1032
         ||||||||||||||||||||||||||||||   ||||||||||||||||
     143 ArgProPheThrLeuIleAlaValAspGlyGluAspHisArgArgLeuAr 159

    1031 CCGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGG 982
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     159 gArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrA 176

     981 ATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCC 932
         ||||||||||||||||||   |||||||||||||||||||||||:::
     176 spArgIleAlaAlaIleAlaAspArgLeuLeuThrGluLeuAlaAspSer 192

     931 TCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCA 882
         |||   |||||||||:::||||||||||||||||||||||||||||||||
     193 SerAspArgSerGlyGluProAlaGluLeuIleGlyGlyPheAlaTyrHi 209

     881 CTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATC 832
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     209 sPheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspP 226

     831 CGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGC 782
         ||||||||||||||||||||:::||||||||||||||||||||||||||
     226 roAlaMetAlaArgGluAlaValGlyValLeuLysAlaLeuGlyLeuGly 242

     781 GGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCC 732
         ||||||||||||:::||||||||||||||||||||||||   |||||
     243 GlyProGlnSerAlaGlyGlyAspGlyThrAspProAlaGlyAspValPr 259

     731 GGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCC 682
         ||||||||||||||||||||||||||||||||||||||||||||:::||||
     259 oAspThrSerAlaLeuGluSerLeuLeuLeuGluAlaValHisAlaAlaA 276

     681 GGCGGAACGACACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAG 632
         |||||:::||||||   ||||||||||||||||||||||||||||||||
     276 rgArgLysAspThrArgThrMetThrArgValLeuTyrGluArgAlaGln 292

     631 GCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGG 582
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     293 AlaGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGl 309

     581 GCTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGC 532
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     309 yLeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeuL 326

     531 TCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCC 482
         ||||||||||||||||||||||||||||||||||||   ||||||
     326 euAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspGlyAspAla 342

     481 GTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTA 432
         :::||||||||||||||||||||||||||:::||||||||||||||||||
     343 IleSerArgPheValGluGluAlaLeuArgHisHisProProValProTy 359

     431 CACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGC 382
         |:::|||||||||||||||||||||||||||:::|||   ||||||||||
     359 rSerLeuTrpArgPheAlaAlaThrGluValValIleArgGlyValArgL 376

     381 TGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGAC 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     376 euProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAsp 392

     331 GGCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTG 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     393 GlyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerAr 409

     281 GCGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGC 232
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     409 gArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnL 426

     231 TCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTC 182
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     426 euAlaGlnLeuGluSerArgThrMetIleGlyValLeuArgSerArgPhe 442

     181 CCCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAA 132
         |||:::|||||||||||||||||||:::||||||||||||||||||||
     443 ProGlnAlaArgLeuAlaValProTyrGluGluLeuArgTrpCysArgLy 459

     131 GGGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGC 86
         |||||||||||||||||||||:::||||||||||||||||||
     459 sGlyAlaGlnThrAlaArgLeuThrAspLeuProValTrpLeuArg 474

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36128

seq_documentation_block:
ID   AAW36128 standard; Protein; 422 AA.
XX
AC   AAW36128;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Daunomycin C-14 hydroxylase.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy.
XX
OS   Streptomyces sp. strain C5.
XX
PN   WO9744439-A2.
```

XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08690.
XX
PR 24-MAY-1996; 96US-0653650.
XX
PA (OHIS) UNIV OHIO STATE. RES FOUND.
XX
PI DeSanti CL, Dickens ML, Strohl WA;
XX
DR WPI; 1998-018495/02.
DR N-PSDB; AAV01447, AAV01451-53.
XX
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT - also hydroxylation and oxidation of other anthracycline(s) with
PT the same enzyme
XX
PS Claim 9; Page 31-33; 59pp; English.

CC This protein comprises the daunomycin C-14 hydroxylase (DC14H)
CC of Streptomyces sp. strain C5. DC14H is a P450-like enzyme useful
CC for the production of doxorubicin from daunomycin, and for the
CC hydroxylation and oxidation of other anthracyclines. Its amino
CC acid sequence was deduced from the doxA gene (see AAV01447)
CC identified in the daunomycin biosynthesis gene cluster from C5 (see
CC AAV01453). Plasmid vectors (see AAV01451-52) have been constructed
CC for use in the recombinant production of DC14H in transformed host
CC cells. Claimed methods for production of doxorubicin include: the
CC use of a host microorganism transformed with a plasmid which
CC contains the doxA gene; incubating a culture of Streptomyces sp.
CC strain C5 with daunomycin; and incubating DC14H with daunomycin.
CC DC14H can also be used in claimed methods of producing
CC 13-dihydrocarminomycin and carminomycin from 13-deoxycarminomycin,
CC and of producing 13-dihydrodaunomycin and daunomycin from
CC 13-deoxydaunomycin. The methods preclude the use of halogens as
CC required in the current chemical methods.
XX
SQ Sequence 422 AA;

alignment_scores:
Quality: 2076.00 Length: 422
Ratio: 5.027 Gaps: 0
Percent Similarity: 97.867 Percent Identity: 94.076

alignment_block:
US-09-673-254-1/rev x AAW36128 ..

Align seg 1/1 to: AAW36128 from: 1 to: 422

```
 351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
     :::|||||||||||||||||||||||||||||||||||:::|||||||
   1 MetSerGlyGluAlaProArgValAlaValAspProPheSerCysProMe 17

1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
     !||||||||||||||||||||||||||||||||||||||||||||||||
  17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50

1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValLy 67

1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCG 1102
     |   |||||||||||||||:::|||||||||||||||||||||||||||||
  67 sGlyProAspLeuAlaProThrAlaTrpArgGlyValAspAspGlyLeuA 84

1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
     |||   ||||||||||||||||||||||||||||||||||||||||||||
 101 GluAspHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117

1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
     ||||||||||||||||||||||||||||||||||||||   |||||||
 117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaAspArgLeuL 134

 951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
     ||||||||||||||||:::|||   |||||||||:::|||||||||||||
 134 euThrGluLeuAlaAspSerSerAspArgSerGlyGluProAlaGluLeu 150

 901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 IleGlyGlyPheAlaTyrHisPheProLeuLeuValIleCysGluLeuLe 167

 851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
     ||||||||||||||||||||||||||||||||||||||||||:::|||||
 167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValGlyValL 184

 801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
     ||||||||||||||||||||||||||||||:::|||||||||||||||||
 184 euLysAlaLeuGlyLeuGlyGlyProGlnSerAlaGlyGlyAspGlyThr 200

 751 GACCCTGCCGGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
     ||||||||||||   ||||||||||||||||||||||||||||||||||
 201 AspProAlaGlyAspValProAspThrSerAlaLeuGluSerLeuLeuLe 217

 701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
     ||||||||||||||:::||||||||||:::||||||   ||||||||||||
 217 uGluAlaValHisAlaAlaArgArgLysAspThrArgThrMetThrArgV 234

 651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250

 601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrGl 267

 551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284

 501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
     ||||||||||   ||||||:::|||||||||||||||||||||||||||
 284 laAspAlaAspGlyAspAlaIleSerArgPheValGluGluAlaLeuArg 300

 451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
     :::|||||||||||||||||:::|||||||||||||||||||||||||||
 301 HisHisProProValProTyrSerLeuTrpArgPheAlaAlaThrGluVa 317

 401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
     |:::|||   ||||||||||||||||||||||||||||||||||||||||
 317 lValIleArgGlyValArgLeuProArgGlyAlaProValLeuValAspI 334

 351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350

 301 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
     ||||||||||||||||   ||||||||||||||||||||||||||||||
 351 HisProAspArgProSerArgArgArgLeuThrPheGlyAspGlyProHi 367

 251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384

 201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
     ||||||||||||||||||||||:::|||||||||||||||||||||:::
 384 lyValLeuArgSerArgProProGlnAlaArgLeuAlaValProTyrGlu 400
```

```
     151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     401 GluLeuArgTrpCysArgLysGlyAlaGlnThrAlaArgLeuThrAspLe 417

     101 GCCCGTCTGGCTGCGC 86
         ||||||||||||||||
     417 uProValTrpLeuArg 422

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36134

seq_documentation_block:
ID   AAW36134 standard; Protein; 275 AA.
XX
AC   AAW36134;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Streptomyces orfA gene product.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy.
XX
OS   Streptomyces sp.
XX
PN   WO9744439-A2.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08690.
XX
PR   24-MAY-1996;   96US-0653650.
XX
PA   (OHIS ) UNIV OHIO STATE. RES FOUND.
XX
PI   DeSanti CL,  Dickens ML,  Strohl WA;
XX
DR   WPI; 1998-018495/02.
DR   N-PSDB; AAV01453.
XX
PT   Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT   - also hydroxylation and oxidation of other anthracycline(s) with
PT   the same enzyme
XX
PS   Disclosure; Fig 3; 59pp; English.

     This protein is the encoded product of the Streptomyces sp. strain
     C5 daunomycin biosynthetic gene cluster orfA (see AAV01453).  This
     gene lies upstream of the doxA (see also AAV01447) gene.  The doxA
CC   gene codes for daunomycin C-14 hydroxylase (see AAW36128), an enzyme
CC   capable of converting daunomycin to the anticancer agent doxorubicin.
CC   Host cells, especially Streptomyces host cells, transformed with
CC   plasmids (see AAV01451-52) containing the doxA gene can be used in
CC   methods for the production of doxorubicin from daunomycin or for the
CC   hydroxylation and oxidation of other anthracyclines.
XX
SQ   Sequence   275 AA;


alignment_scores:
            Quality: 1385.00              Length:    275
              Ratio:  5.149                Gaps:      0
 Percent Similarity:  97.818   Percent Identity: 94.182

alignment_block:
 US-09-673-254-1/rev x AAW36134    ..

 Align seg 1/1  to: AAW36134  from: 1  to: 275

    2175 GTGACCAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCC 2126
         :::|||||||||||||||||||||||||||||||||||||||||||||||
       1 MetThrArgPheAlaProGlyAlaProAlaTrpPheAspLeuGlySerPr 17

    2125 CGATGTCGCCGCCTCGGCCGACTTCTACACCGGCCTGTTCGGCTGGACCG 2076
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 oAspValAlaAlaSerAlaAspPheTyrThrGlyLeuPheGlyTrpThrA 34

    2075 CCACCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACGTTCAGCTCC 2026
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 laThrValValSerAspProGlyAlaGlyGlyTyrThrThrPheSerSer 50

    2025 GACGGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCCCTA 1976
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AspGlyLysLeuValAlaAlaValAlaArgHisGlnIleAspThrProTy 67

    1975 CCACCGGCCGTACGGGCCCGGGAACGACCAGCACGGCATGCCGGCCATCT 1926
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rHisArgProTyrGlyProGlyAsnAspGlnHisGlyMetProAlaIleT 84

    1925 GGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGGGTC 1876
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
      84 rpThrValTyrPheAlaThrAsnAspAlaAspAlaLeuThrLysArgVal 100

    1875 GAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGGCCT 1826
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
     101 GluThrAlaGlyGlyAspValIleMetThrProMetAspValLeuGlyLe 117

    1825 CGGCCGGATGGCGGTCTTCGCCGACCCCGCCGGGGCCGCGTTCGCGGTCT 1776
         ||||||||||||||||||||||||||:::|||||||||||||||||||||
     117 uGlyArgMetAlaValPheAlaAspProSerGlyAlaAlaPheAlaValT 134

    1775 GGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGCTCG 1726
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rpArgLysGlyValMetGluGlyAlaGluValThrGlyValProGlySer 150

    1725 GTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCGGGACTT 1676
         |||||||||||||||||||||||||||    ||||||:::||||||   ||
     151 ValGlyTrpValGluLeuValThrAspAspIleGlyThrAlaArgGlyPh 167

    1675 CTACCCGGCGACCCTCGGCCTGGCTCCGGCCGACACCGGACTGAAGGGCG 1626
         ||||   ||||||||||||||||||||||||||||||   |||||||
     167 eTyrArgAlaThrLeuGlyLeuAlaProAlaAspThrGlyArgLysGlyV 184

    1625 TCACCGACCCGGTCTGGCACATCGGTGACACACCGGTCGCCGGCACCCAG 1576
         ||||||||||||||||||||   ||||||||||||||||||||:::
     184 alThrAspProValTrpHisIleHisAspThrProValAlaGlyThrArg 200

    1575 GAGCTGGGCGTCACCGGCGCGGTACGGCCGCACTGGGCCGTGCTGTTCGC 1526
         |||||||||:::||||||||||||||||||||||||||||||||||||::
     201 GluLeuGlyThrThrGlyAlaValArgProHisTrpAlaValLeuPheSe 217

    1525 CGTGCACGACTGCGACGCGACGGTCCGGCGCGCCGTTGAACTCGGCGGCT 1476
         :||||||||||||||||||||||||||||||||||||||||||||||||||
     217 rValHisAspCysAspAlaThrValArgArgAlaValGluLeuGlyGlyS 234

    1475 CCGTCGAGAACGAGCCCGCCGACACGCCCAGGGGGCGGCGGGCGGACCTG 1426
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     234 erValGluAsnGluProValAspThrProArgGlyArgArgAlaAspLeu 250

    1425 CTCGACCCGCACGGGGCCGGCTTCTCGGTGGTCGAACTGCGGGAGGGGTA 1376
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     251 LeuAspProHisGlyAlaGlyPheSerValValGluLeuArgGluAlaTy 267

    1375 CCCCGCGGCGGCGGGCGGTGCCTCG 1351
         |||||||||||||   |||||||||
     267 rProAlaAlaAlaAspGlyAlaSer 275

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW68559

seq_documentation_block:
ID   AAW68559 standard; Protein; 287 AA.
XX
AC   AAW68559;
XX
DT   25-JAN-1999  (first entry)
```

```
XX
DE   S. peucetius dnrU gene product.
XX
KW   dnrU; dnrX; recombinant; microorganism; doxorubicin; daunorubicin;
KW   metabolism; baumycin; secretion; anti-tumour compound.
XX
OS   Streptococcus peucetius.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 59
FT                   /note= "encoded by AAC"
FT   Misc-difference 124
FT                   /note= "encoded by AAT"
FT   Misc-difference 125
FT                   /note= "encoded by CAC"
FT   Misc-difference 167
FT                   /note= "encoded by CAC"
FT   Misc-difference 184
FT                   /note= "encoded by ATG"
     Misc-difference 202
                     /note= "encoded by CAC"

     WO9839458-A1.
XX
PD   11-SEP-1998.
XX
PF   05-MAR-1998;   98WO-US03938.
XX
PR   28-JUL-1997;   97US-0901306.
PR   06-MAR-1997;   97US-0812412.
XX
PA   (PHAA ) PHARMACIA & UPJOHN SPA.
XX
PI   Colombo AL,  Filippini S,  Fonstein L,  Htutchinson RC;
PI   Lomovskaya N;
XX
DR   WPI; 1998-506367/43.
DR   N-PSDB; AAV34001.
XX
PT   Production of doxorubicin and daunorubicin - using a recombinant
PT   daunorubicin producing microorganism in which at least one dnrU or
PT   dnrX metabolism gene has been inactivated
XX
PS   Disclosure; Fig 2A-B; 48pp; English.
XX
CC   This sequence represents the dnrU gene product from the microorganism
CC   Streptomyces peucetius.  The invention relates to a recombinant
CC   microorganism which produces doxorubicin (DoR), in which at least 1
CC   daunorubicin (DaR) metabolism gene selected from dnrU and dnrX, has been
CC   inactivated.  The microorganism is preferably S. peucetius WMH1658,
     WMH1654 and WMH1662.  The function of the DaR gene is inactivated by
     insertion of a gene (preferably neomycin/kanamycin resistance gene aphII)
     into the daunorubicin metabolism gene.  The blocking of the function of
     at least one gene of DaR metabolism can increase DaR and DoR production
     levels and cause the disappearance of baumycin-like products resulting
CC   in DaR and DoR secretion directly into the culture medium.  The products
CC   can be used in the production of anti-tumour compounds.
XX
SQ   Sequence   287 AA;


alignment_scores:
            Quality:  978.00              Length:    196
              Ratio:  5.041                 Gaps:      0
 Percent Similarity:  98.980   Percent Identity: 97.449

alignment_block:
 US-09-673-254-1/rev x AAW68559    ..

 Align seg 1/1  to: AAW68559  from: 1  to: 287

    2869 CGGTACCCGCGCATCGATGTCATGGCCAGCAACGCCGGCGGCATGTTCTG 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||

      92 ArgTyrProArgIleAspValMetAlaSerAsnAlaGlyGlyMetPheTr 108

    2819 GTCGCGCACCACGACCCAGGACGGGTTCGAGGCCACCATCCAGGTCAATC 2770
         ||||||||||||||||||||||||||||||||||||||||||||||||   :
     108 pSerArgThrThrThrGlnAspGlyPheGluAlaThrIleGlnValMetA 125

    2769 ACCTCGCAGGCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGC 2720
         ::|||||||||||||||||||||||||||||||||||||||||||||||
     125 snLeuAlaGlyPheLeuLeuAlaArgLeuLeuArgGluArgLeuAlaGly 141

    2719 GGGCGGCTGATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGGATCGA 2670
         ||||||||||||||||||||||||||||||||||||||||||||||||
     142 GlyArgLeuIleLeuThrSerSerAspAlaTyrThrGlnGlyArgIleAs 158

    2669 CCCGGACGACCTCAACGGCGACCGTCACCGCTACAGCGCCGGCCAGGCGT 2620
         ||||||||||||||||||||||||:::||||||||||||||||||||||
     158 pProAspAspLeuAsnGlyAspArgAsnArgTyrSerAlaGlyGlnAlaT 175

    2619 ACGGCACGTCCAAACAGGCCAACATCATGACCGCGGCGGAGGCCGCCAGG 2570
         ||||||||||||||||||||||||   ||||||||||||||||||||||
     175 yrGlyThrSerLysGlnAlaAsnIleAsnThrAlaAlaGluAlaAlaArg 191

    2569 CGCTGGCCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTCCGCAC 2520
         ||||||||||||||||||||||||||||:::||||||||||||||||||
     192 ArgTrpProAspValLeuAlaValSerTyrAsnProGlyGluValArgTh 208

    2519 CCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCAACCCCT 2470
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     208 rArgIleGlyArgGlyThrValAlaSerSerTyrPheArgPheAsnProP 225

    2469 TCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTGGCGTCC 2420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     225 heLeuArgSerAlaAlaLysGlyAlaAspThrLeuValTrpLeuAlaSer 241

    2419 GCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGACCGGCGGCT 2370
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     242 AlaProAlaGluGluLeuThrThrGlyGlyTyrTyrSerAspArgArgLe 258

    2369 GTCCCCGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCT 2320
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     258 uSerProValSerGlyProThrAlaAspAlaGlyLeuAlaAlaLysLeuT 275

    2319 GGGAGGCCGGCGCGGCCGCCGTCGGCGACACCGCGCAC 2282
         |||||||||||||||||||||||||||||||||||||
     275 rpGluAlaGlyAlaAlaAlaValGlyAspThrAlaHis 287

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW36133

seq_documentation_block:
ID   AAW36133 standard; Protein; 193 AA.
XX
AC   AAW36133;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Streptomyces orf1 gene product.
XX
KW   Daunomycin C-14 hydroxylase; doxA gene; doxorubicin; daunomycin;
KW   13-dihydrocarminomycin; carminomycin; anthracycline;
KW   anticancer; cytostatic; cancer; therapy.
XX
OS   Streptomyces sp.
XX
PN   WO9744439-A2.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08690.
XX
PR   24-MAY-1996;   96US-0653650.
XX
PA   (OHIS ) UNIV OHIO STATE. RES FOUND.
XX
```

PI DeSanti CL, Dickens ML, Strohl WA;
XX
DR WPI; 1998-018495/02.
DR N-PSDB; AAV01453.
XX
PT Producing doxorubicin from daunomycin by daunomycin C14 hydroxylase
PT - also hydroxylation and oxidation of other anthracycline(s) with
PT the same enzyme
XX
PS Disclosure; Fig 3; 59pp; English.
XX
CC This protein is the encoded product of the Streptomyces sp. strain
CC C5 daunomycin biosynthetic gene cluster orf1 (see AAV01453). This
CC gene lies upstream of orfA and doxA (see also AAV01447). The doxA
CC gene codes for daunomycin C-14 hydroxylase (see AAW36128), an enzyme
CC capable of converting daunomycin to the anticancer agent doxorubicin.
CC Host cells, especially Streptomyces host cells, transformed with
CC plasmids (see AAV01451-52) containing the doxA gene can be used in
CC methods for the production of doxorubicin from daunomycin or for the
CC hydroxylation and oxidation of other anthracyclines.
XX
SQ Sequence 193 AA;

alignment_scores:
Quality: 961.00 Length: 192
Ratio: 5.005 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.354

alignment_block:
US-09-673-254-1/rev x AAW36133 ..

Align seg 1/1 to: AAW36133 from: 1 to: 193

```
2860 CGCATCGATGTCATGGCCAGCAACGCCGGCGGCATGTTCTGGTCGCGCAC 2811
     ||||||||||||||||||:::||||||||||||||||||||||||||||||
   1 ArgIleAspValMetAlaGlyAsnAlaGlyGlyMetPheTrpSerArgTh 17

2810 CACGACCCAGGACGGGTTCGAGGCCACCATCCAGGTCAATCACCTCGCAG 2761
     ||||||||||||||||||||||||||||:::||||||||||||||||||||
  17 rThrThrGlnAspGlyPheGluAlaThrLeuGlnValAsnHisLeuAlaG 34

2760 GCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGCGGGCGGCTG 2711
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  34 lyPheLeuLeuAlaArgLeuLeuArgGluArgLeuAlaGlyGlyArgLeu 50

2710 ATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGGATCGACCCGGACGA 2661
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 IleLeuThrSerSerAspAlaTyrThrGlnGlyArgIleAspProAspAs 67

2660 CCTCAACGGCGACCGTCACCGCTACAGCGCCGGCCAGGCGTACGGCACGT 2611
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 pLeuAsnGlyAspArgHisArgTyrSerAlaGlyGlnAlaTyrGlyThrS 84

2610 CCAAACAGGCCAACATCATGACCGCGGCGGAGGCCGCCAGGCGCTGGCCG 2561
     ||||||||||||||||||||||||||:::||||||||||||||||||||||
  84 erLysGlnAlaAsnIleMetThrAlaThrGluAlaAlaArgArgTrpPro 100

2560 GACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTCCGCACCCGCATCGG 2511
     |||||||||:::||||||||||||||||||||||||||||||||||||||
 101 AspValLeuThrValSerTyrHisProGlyGluValArgThrArgIleGl 117

2510 ACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCAACCCCTTCCTGCGCT 2461
     ||||||||||||||||||:::||||||||||||||||||||||||||||||
 117 yArgGlyThrValAlaSerThrTyrPheArgPheAsnProPheLeuArgS 134

2460 CCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTGGCGTCCGCGCCGGCC 2411
     ||||||||||||||||||||||||||||||||||||:::|||||||||||
 134 erAlaAlaLysGlyAlaAspThrLeuValTrpLeuAlaAlaAlaProAla 150

2410 GAGGAGTTGACCACGGGCGGCTACTACAGCGACCGGCGGCTGTCCCCGGT 2361
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 GluGluLeuThrThrGlyGlyTyrTyrSerAspArgArgLeuSerProVa 167

2360 GAGCGGCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCTGGGAGGCCG 2311
     ||||||||||||||||||||||||||||||||||||||||||||||||:
  167 lSerGlyProThrAlaAspAlaGlyLeuAlaAlaLysLeuTrpGluAlaS 184

2310 GCGCGGCCGCCGTCGGCGACACCGCG 2285
     ::|||||||||||||||||||||||
  184 erAlaAlaAlaValGlyAspThrAla 192
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW68560

seq_documentation_block:
ID AAW68560 standard; Protein; 117 AA.
XX
AC AAW68560;
XX
DT 25-JAN-1999 (first entry)
XX
DE S. peucetius partial dnrV gene product.
XX
KW dnrU; dnrX; recombinant; microorganism; doxorubicin; daunorubicin;
KW metabolism; baumycin; secretion; anti-tumour compound.
XX
OS Streptococcus peucetius.
XX
FH Key Location/Qualifiers
FT Misc-difference 13
FT /note= "encoded by GAG"
FT Misc-difference 14
FT /note= "encoded by CTC"
FT Misc-difference 68
FT /note= "encoded by CAC"
FT Misc-difference 78
FT /note= "encoded by CAC"
FT Misc-difference 80
FT /note= "encoded by ATG"
FT Misc-difference 88
FT /note= "encoded by TTC"
XX
PN WO9839458-A1.
XX
PD 11-SEP-1998.
XX
PF 05-MAR-1998; 98WO-US03938.
XX
PR 28-JUL-1997; 97US-0901306.
PR 06-MAR-1997; 97US-0812412.
XX
PA (PHAA) PHARMACIA & UPJOHN SPA.
XX
PI Colombo AL, Filippini S, Fonstein L, Htutchinson RC;
PI Lomovskaya N;
XX
DR WPI; 1998-506367/43.
DR N-PSDB; AAV34001.
XX
PT Production of doxorubicin and daunorubicin - using a recombinant
PT daunorubicin producing microorganism in which at least one dnrU or
PT dnrX metabolism gene has been inactivated
XX
PS Disclosure; Fig 2A-B; 48pp; English.
XX
CC This sequence represents a partial dnrV gene product from the
CC microorganism Streptomyces peucetius. The invention relates to a
CC recombinant microorganism which produces doxorubicin (DoR), in which
CC at least 1 daunorubicin (DaR) metabolism gene selected from dnrU and
CC dnrX, has been inactivated. The microorganism is preferably S. peucetius
CC WMH1658, WMH1654 and WMH1662. The function of the DaR gene is inactivated
CC by insertion of a gene (preferably neomycin/kanamycin resistance gene
CC aphII) into the daunorubicin metabolism gene. The blocking of the
CC function of at least one gene of DaR metabolism can increase DaR and DoR
CC production levels and cause the disappearance of baumycin-like products
CC resulting in DaR and DoR secretion directly into the culture medium.

CC The products can be used in the production of anti-tumour compounds.
XX
SQ Sequence 117 AA;

alignment_scores:
Quality: 593.00 Length: 117
Ratio: 5.202 Gaps: 0
Percent Similarity: 97.436 Percent Identity: 95.726

alignment_block:
US-09-673-254-1/rev x AAW68560 ..

Align seg 1/1 to: AAW68560 from: 1 to: 117

```
2175 GTGACCAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCC 2126
     ||||||||||||||||||||||||||||||||||||||||||   ||||||||
   1 ValThrArgPheAlaProGlyAlaProAlaTrpPheAspGlyGlySerPr 17

2125 CGATGTCGCCGCCTCGGCCGACTTCTACACCGGCCTGTTCGGCTGGACCG 2076
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  17 oAspValAlaAlaSerAlaAspPheTyrThrGlyLeuPheGlyTrpThrA 34

2075 CCACCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACGTTCAGCTCC 2026
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  34 laThrValValSerAspProGlyAlaGlyGlyTyrThrThrPheSerSer 50

2025 GACGGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCCCTA 1976
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 AspGlyLysLeuValAlaAlaValAlaArgHisGlnIleAspThrProTy 67

1975 CCACCGGCCGTACGGGCCCGGGAACGACCAGCACGGCATGCCGGCCATCT 1926
     |:::|||||||||||||||||||||||||||:::|||   ||||||||||
  67 rAsnArgProTyrGlyProGlyAsnAspGlnAsnGlyAsnProAlaIleT 84

1925 GGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGGGTC 1876
     ||||||||||||   ||||||||||||||||||||||||||||||||||
  84 rpThrValTyrProAlaThrAspAspAlaAspAlaLeuThrLysArgVal 100

1875 GAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGGCCT 1826
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 GluThrAlaGlyGlyGluValIleMetThrProMetAspValLeuGlyLe 117

1825 C 1825
     |
 117 u 117
```

OM of: US-09-673-254-1 to: PIR_71:* out_format : pfs

Date: Jun 11, 2002 8:10 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09673254/runat_11062002_114211_938/app_query.fasta_1.2973
-DB=PIR_71 -QFMT=fastan -SUFFIX=n2p.rpr -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
[redacted]ANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
[redacted]_MAX=100 -THR_MIN=0 -ALIGN=7 -MODE=LOCAL -OUTFMT=pfs
[redacted]M=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
[redacted]SER=US09673254_@CGN1_1_200 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-673-254-1
Query length: 2870
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 86.320000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
pir2:JQ0405	-	625.50	494.99	1.6e-20	1106	! hypothetical 119.5K protein (u
pir2:JQ0405	+	560.00	443.28	1.2e-17	1106	! hypothetical 119.5K protein (u
pir2:S27923	-	538.00	427.37	1.1e-16	924	! gene LF3 protein - human herpes
pir2:T02345	-	486.50	381.33	2.1e-14	1791	! hypothetical protein KIAA0324
pir2:T02345	+	479.00	375.41	4.4e-14	1791	! hypothetical protein KIAA0324
pir1:B45344	+	468.50	367.39	1.3e-13	1733	! probable nuclear antigen - sui
pir1:A35175	-	464.50	366.29	1.9e-13	1344	! mucin 1 precursor, repetitive
pir1:CGCH1S	-	459.50	364.41	3.1e-13	1042	! collagen alpha 1(I) chain - ch
pir1:CGHU1S	-	458.00	360.47	3.7e-13	1464	! collagen alpha 1(I) chain prec
pir2:T34434	-	454.00	353.89	5.6e-13	2232	! hypothetical protein K06A9.1a
pir2:S21626	-	450.00	354.21	8.3e-13	1453	! collagen alpha 1(I) chain prec
pir2:S59856	-	450.00	354.15	8.3e-13	1464	! collagen alpha 1(III) chain pr
pir2:S27923	+	448.50	356.70	9.5e-13	924	! gene LF3 protein - human herpes
[redacted]:QQBE3	-	442.00	354.30	1.8e-12	660	! BHLF1 protein - human herpesvir
[redacted]:CGBO7S	-	442.00	350.54	1.8e-12	1049	! collagen alpha 1(III) chain -
[redacted]:CGHU2V	-	439.00	345.29	2.5e-12	1496	! collagen alpha 2(V) chain prec
[redacted]:B45344	-	437.50	342.91	3.0e-12	1733	! probable nuclear antigen - sui
pir2:H70933	-	434.50	355.91	3.7e-12	261	! hypothetical protein Rv0577 - M
pir2:B40505	+	434.50	339.55	4.0e-12	1958	! hypothetical protein - suid he
pir1:CGHU7L	-	432.50	340.32	4.9e-12	1466	! collagen alpha 1(III) chain pr
pir2:T43481	-	430.50	346.27	5.8e-12	580	! probable mucin DKFZp434C196.1 -
pir2:A43932	-	430.50	332.88	6.2e-12	3020	! mucin 2 precursor, intestinal
pir2:T35032	-	429.00	351.50	6.5e-12	263	! probable hydroxylase - Streptom
pir2:T45467	-	428.00	337.04	7.7e-12	1418	! collagen alpha 1(II) chain pre
pir2:B40505	-	427.50	334.03	8.2e-12	1958	! hypothetical protein - suid he
pir2:A41182	-	425.50	335.06	9.9e-12	1419	! collagen alpha 1(II) chain pre
pir2:T18535	-	419.00	331.63	1.9e-11	1151	! high molecular mass nuclear ar
pir2:B41182	-	419.00	329.55	1.9e-11	1487	! collagen alpha 1(II) chain pre
pir2:S18803	+	419.00	327.81	1.9e-11	1843	! collagen alpha 1(V) chain - ha
pir2:I49607	-	418.00	328.71	2.1e-11	1497	! procollagen type V alpha 2 - m
pir1:CGHU1V	+	418.00	327.04	2.1e-11	1838	! collagen alpha 1(V) chain prec
pir1:S23809	-	415.00	326.80	2.9e-11	1414	! collagen alpha 2(I) chain homo
pir2:A54849	+	414.50	320.45	3.1e-11	2944	! collagen alpha 1(VII) chain pr
pir1:CGHU1S	+	413.50	325.33	3.3e-11	1464	! collagen alpha 1(I) chain prec
pir1:CGHU7L	+	413.50	325.32	3.3e-11	1466	! collagen alpha 1(III) chain pr
pir1:CGHU6C	-	412.00	324.02	3.9e-11	1487	! collagen alpha 1(II) chain pre
pir1:B42606	-	411.50	334.18	3.9e-11	405	! cytochrome P450 CVIIB1 - Saccha
pir1:CGHU2S	-	410.50	323.53	4.5e-11	1366	! collagen alpha 2(I) chain prec
pir2:T35325	-	410.00	336.44	4.5e-11	265	! probable hydroxylase - Streptom
pir1:S48478	-	408.00	321.55	5.8e-11	1367	! glucan 1,4-alpha-glucosidase (
pir2:T45025	-	408.00	313.76	6.1e-11	3570	! mucin MUC5B, tracheobronchi
pir2:S24750	-	406.50	330.09	6.5e-11	412	! cytochrome P450 (soy) - Stre
pir2:F75518	-	405.50	323.53	7.4e-11	839	! hypothetical protein - Deino
pir2:A40333	-	405.50	318.86	7.5e-11	1492	! collagen alpha 1'(II) chain
pir2:S59856	+	405.00	318.62	7.9e-11	1464	! collagen alpha 1(III) chain

seq_name: pir2:JQ0405

seq_documentation_block:
hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
N;Alternate names: ORF 1 protein
C;Species: Micrococcus luteus, Micrococcus lysodeikticus
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C;Accession: JQ0405
R;Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A;Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification
A;Reference number: S04781; MUID:89364717
A;Accession: JQ0405
A;Molecule type: DNA
A;Residues: 1-1106 <SHI>
A;Cross-references: EMBL:X15867
A;Note: this reading frame extends between two stop codons and does not begin with a
A;Note: the gene encoding this protein overlaps uvrA gene
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

alignment_scores:
Quality: 625.50 Length: 1215
Ratio: 1.295 Gaps: 69
Percent Similarity: 39.753 Percent Identity: 29.465

alignment_block:
US-09-673-254-1/rev x JQ0405 ..

Align seg 1/1 to: JQ0405 from: 1 to: 1106

```
2813 CACCACGACCCAGGACGGGTTCGAGGCCACCATCCAGGTCAATCACCTCG 2764
     ||||||         |||  ||||||     |||   :::   |||||
  60 HisHisArgLeuLeuArgArgArgGly.....ProArgArgArgProAr 74

2763 CAGGCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGCGGGCGG 2714
     |      |||         |||                  |||||||||:
  74 gLeuArgProArgArgArgAlaLeu.............GlyArgAlaG 86

2713 CTGATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGGATCGACCCGGA 2664
     ::   ||||||                           |||||||||
  86 ly...ProHis.........................ArgArgProGly 92

2663 CGACCTCAACGGCGACCGTCACCGCTACAGCGCCGGCCAGGCGTACGGCA 2614
     |||||||::   :::|||
  93 ArgProArgAlaGlnPro.............................. 98

2613 CGTCCAAACAGGCCAACATCATGACCGCGGCGGAGGCCGCCAGGCGCTGG 2564
                              :::|||   |||   |||      |
  99 .......................GluGlyArgGlyCysGlnLeuProA 107

2563 CCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTCCGCACCCGCAT 2514
     ||||||||                                      |||
 107 laGlyArg.....................................His 110

2513 CGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCAACCCCTTCCTGC 2464
           ||||||   :::   |||         |||   |||||::
 111 GlyArgValHisGlySerValArgLeu......GlyGlnValLeuProGl 125

2463 GCTCCGCGGCGAAGGGCGCCGACACCCT.............CGTGTGGCT. 2427
     :||||||   :::   ||||||   |||             |||||||||
 125 yLeuArgHisAspLeuArgArgGlyProAlaAlaLeuArgArgValAlaL 142

2426 ............................GGCGTCCGCGCCGGCCGAGGA 2406
                                 |||   :::||||||||||||
 142 euLeuLeuArgProHisValProGlyProGlyGlyGlnAlaGlyArgGly 158
```

```
 648 gArgArgAlaArgArgGlyHisAspArgArgGlyGlyLeuAspArgGlyH 665

 962 ..CGGCCGGCTGCTCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAA 915
       |||||||:::      |||:::||||||||   |||   :::|||
 665 isArgProSerArgGlyArgValArgArgArgGlyArgAlaLeuGlyLeu 681

 914 ACCGGC..............CGAGCTGATCGGCGGCTTCGCGTACCACT 880
       |||                :::|||   |||:::   |||:::|||||
 682 ProGlyGlySerGlnGlyGluHisAlaValArgHisArgArgLeuProLe 698

 879 TCCCGCTGTTGGTCATCTGCGAGCT..................... 855
     |           |||      |||
 698 uArgProProLeuHisArgGlyAlaGlyAlaAlaSerArgProGlyGluG 715

 854 ..GCTCGGTGTGCCGGT...CACCGATCCGGCGATGGCCCGCGAGGCCGT 810
       ||||||:::   |||   :::|||   |||:::   |||:::||||||
 715 lyAlaArgAlaAspGlyProArgArgProGlyGluGlnProGluGlyArg 731

 810 ..................................................... 810

 732 LeuGlyProGlyProAlaArgGlyProHisGlyArgAspGlyArgValAr 748

 809 ....................CAGCGTTCTCAAGGCACTCGGCCTCGGC 782
                         |||||||:::            |||   :
 748 gLeuArgGlnValHisAlaAspGlnArgAspProLeuGlnGlyProGlyG 765

 781 GGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCC 732
     ::||||||:::||||||      |||      |||:::      |||:::
 765 lnProAlaGlnArgArgGlnAlaArgAlaArgProAlaProValArgGly 781

 731 GGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGC.............. 696
     |||      :::|||||||:::   :::||||||||:::
 782 GlyThrArgAlaProGlyGlnGlyGlyProArgGlyProGluProHisAr 798

 695 .......CGTGCACTCAGCCCGGCGGAACGACACCCC.............. 666
               :::|||:::|||   :::   |||
 798 gAlaHisAlaThrLeuGlnProArgHisLeuHisGlyArgValArgArgA 815

 666 ..................................................... 666

 815 spProGlnAlaLeuArgGlyAspProArgGlyGluGlyProGlyLeuPro 831

 665 ..................GACCATGACCCGCGTGCTGTACGAGCGCGCGC 634
                        :::|||:::   |||||||:::|||:::
 832 AlaGlyProValLeuLeuGlnHisGlnGlyArgAlaLeuArgGlyValAr 848

 633 AGGCCGA.......................................... 627
      ||||||
 848 gGlyArgArgHisAlaGluAspArgAspGluLeuProAlaGlyArgLeuA 865

 626 .....GTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGG 582
           ||||||          |||   |||::::::       |||   |||
 865 rgAlaValArgGlyValProArgGlyProValGlnProGlyAspAlaArg 881

 581 GCTCATCTT.................CGCCGGCCACGACAC........ 558
     :::|||||| .                |||   |||||||||
 882 GlyHisLeuGlnGlyGlnGluHisArgArgGlyProArgHisAlaAspAr 898

 557 ................................CACCGGCTCCTTCCTG 542
                                     |||      :::||||
 898 gGlyGlyArgGlyLeuLeuGlnArgValHisProHisLeuAlaValProG 915

 541 GGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGA 492
     ||         |||         |||       |||   |||   :::|||
 915 lyHisAlaArgArgArgSerGlyLeuArgProSerGlyProAlaArg 931

 491 CGA.........GGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGCT 451
     :::          |||||||||   |||:::|||   :::|||:::
 932 HisHisAlaLeuGlyArgArgGlyProAlaArgGluAlaGlyGlyArgAl 948

 450 ACCACCCGCC......GGTGCCCTACACGTTGTGGAGGTTCGCTGC...C 410
      |||   |||               |||:::         |||:::   |
 948 aProGluAlaLeuGlnArgProHisHisLeuArgAlaGlyArgAlaAspH 965
```

```
 409 ACGGAGGTGACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCC...... 366
     |||||         ||||||:::|||   ||||||      |||
 965 isGlyValAlaLeuArgArgHisProGlnAlaProAlaArgAlaSerVal 981

 365 ......GGTGCTGGTGGACATCGAGGGCACCAACACCGACGGCCGCCATC 322
           |||   |||   |||   :::|||   ||||||      |||
 982 ProArgGlyGlnGlyGlnHisGlyAlaHis...HisArgAlaGlnProAr 997

 321 ACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTGGCGGCGGCTC 272
      |||          |||  |||    |||||| ||||||  ||||||  :
 997 gArgAspGlnGluArgGlyPro.ArgAspArg.ProArgProGlyGly.A 1013

 271 ACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGCTCGCCCAGCT 222
     ::|||||||||||||               ||||||   ::::::   :::
1013 rgLeuArgArgArgHis.........AspArgGlyHisGlyHisAlaGly 1026

 221 GGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTCCCCGAGGCCC 172
     |||                  |||:::   ::::::   |||:::|||||
1027 Gly..............GlyArgAlaArgArgArgGluProHisGlyPr 1038

 171 GACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCG.............. 135
     |:::   |||::::::   |||||||||   :::|||
1038 oValProArgGlyAlaAlaArgValAlaTrpThrProProThrGlnAspG 1055

 134 .....GAAGGGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCT 90
          ||||||   |||
1055 luAspGluGlyGluPro................................ 1060

  89 GCGCTGATCCGCCGGCCGCGACCGGCACGGGACCGCCCACCGCGC...AC 43
           ||||||   |||||||::       |||||||||
1061 ......TrpArgArgArgArgProCysAlaGlyAlaProProArgCysCy 1075

  42 CGCGCGGTGGGCGGTCCCGTGCCGTGTACCCGGTGCG 6
     :::      |||||||||           ||||||
1075 sGlySerProAlaValProArgSerSerAspGlyAla 1087
```

seq_name: pir2:JQ0405

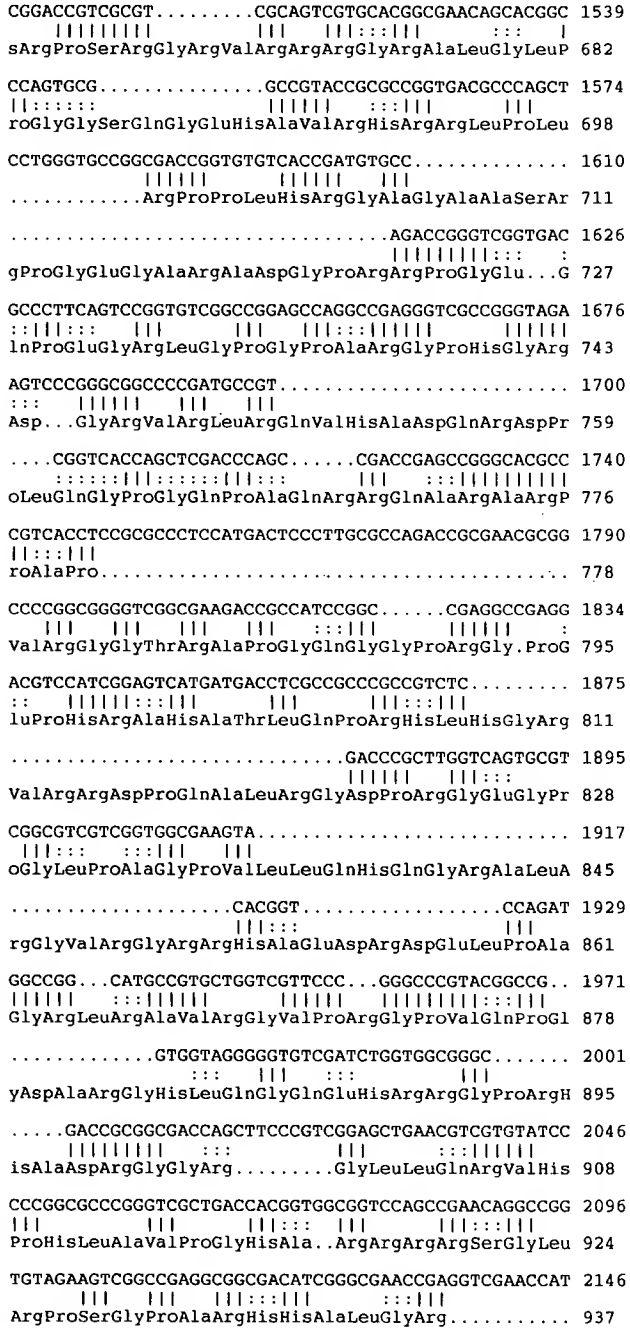
seq_documentation_block:
hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
N;Alternate names: ORF 1 protein
C;Species: Micrococcus luteus, Micrococcus lysodeikticus
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C;Accession: JQ0405
R;Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A;Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification
A;Reference number: S04781; MUID:89364717
A;Accession: JQ0405
A;Molecule type: DNA
A;Residues: 1-1106 <SHI>
A;Cross-references: EMBL:X15867
A;Note: this reading frame extends between two stop codons and does not begin with a
A;Note: the gene encoding this protein overlaps uvrA gene
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

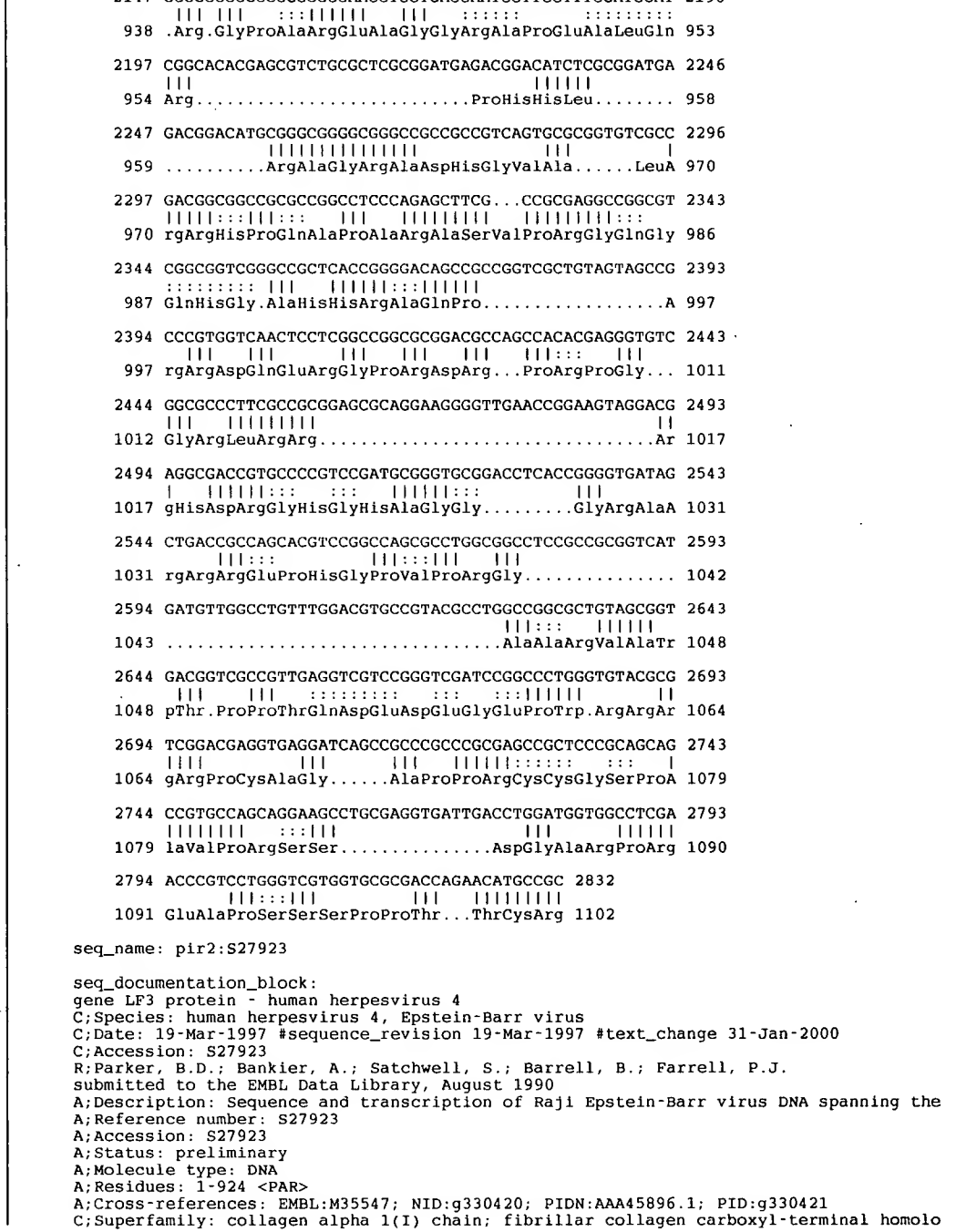
alignment_scores:
Quality: 560.00 Length: 1168
Ratio: 1.131 Gaps: 78
Percent Similarity: 42.380 Percent Identity: 30.908

alignment_block:
US-09-673-254-1 x JQ0405 ..

Align seg 1/1 to: JQ0405 from: 1 to: 1106

```
  10 CCGGGTACACGGCACGGGACCGCCCACCGCGCGGTGCGCGGTGGGCGGTC 59
     |||:::   |||||||||   :::|||   :::||||||   |||:::
 106 ProAlaGlyArgHisGlyArgValHisGlySerValArgLeuGlyGlnVa 122

  60 CCGTGCCGGT.................CGCGGC............CGGC 79
```





```
seq_name: pir2:S27923

seq_documentation_block:
gene LF3 protein - human herpesvirus 4
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Jan-2000
C;Accession: S27923
R;Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.; Farrell, P.J.
submitted to the EMBL Data Library, August 1990
A;Description: Sequence and transcription of Raji Epstein-Barr virus DNA spanning the
A;Reference number: S27923
A;Accession: S27923
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-924 <PAR>
A;Cross-references: EMBL:M35547; NID:g330420; PIDN:AAA45896.1; PID:g330421
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
```

```
alignment_scores:
            Quality:  538.00              Length:   1034
              Ratio:  1.231                Gaps:     63
 Percent Similarity:  42.263   Percent Identity: 30.077

alignment_block:
 US-09-673-254-1/rev x S27923   ..

 Align seg 1/1  to: S27923  from: 1  to: 924

    2804 CCAGGACGGGTTCGAGGCCACCATCCAGGTCAATCACCTCGCAGGCTTCC 2755
      12 ProGlySerGlyLeuGlyAlaHisPro......HisProArgArg..... 24

    2754 TGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGCGGGCGGCTGATCCTC 2705
      25 .SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProA 41

    2704 ACCTCGTCCGACGCGTACACCCAGGGCCGGAT.............CGA 2670
      41 rgAla..............ProGlyProGluProArgThrArgLeuGln 52

    2669 CCCGGACGACCTCAACGG......CGACCGTCACCGCTACAGCGCCGGCC 2626
      53 ProAlaThrProArgArgSerGlyAlaAlaAspProAlaAsp..ProVal 68

    2625 AGGCGTACGGCACGTCCAAACAGGCCAACATCATGACCGCGGCGGAGGCC 2576
      69 GlyHisProAlaAlaProArgAlaProGlyProGluProArgThrArgLe 85

    2575 GCCAGGCGCTGGCCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGT 2526
      85 uGlnProAla.......................ThrProArgArgS 93

    2525 CCGCACCCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCA 2476
      93 er..............GlyAlaAlaAspProAlaAspProValGlyHis 104

    2475 ACCCCTTCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTG 2426
     105 ......ProAlaAlaProArgAlaProGlyProGluProArgThrArgLe 119

    2425 GCGTCCGCGCCGGCCGAGG............AGTTGACCACGG....... 2395
     119 uGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProV 136

    2394 ..........GCGGCTACTACAGCGACCGGCGGCTGTCCCCGGTGAGCG 2356
     136 alGlyHisProAlaAlaProArgAlaProGlyProGluProArgThrArg 152

    2355 GCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCT.............. 2320
     153 LeuGlnProAlaThrPro.....ArgArgSerGlyAlaAlaAspProAl 167

    2319 ..........GGGAGGCCGGCGCGGCCGCCGTCG............... 2296
     167 aAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProA 184

    2295 ................GCGACACCGCGCACTGACGGCGGCGGCCCGCCC 2263
     184 rgThrArgLeuGlnProAlaThrProArgArgSerGly.AlaAlaAspPr 200

    2262 CGCCCGCATGTCCGTCTCATCCGCGAGATGTCCGTCTCATCCGCGAGCGC 2213
     200 oAla............AspProValGlyHisProAlaAlaProArgAla. 212

    2212 AGACGCTCGTGTGCCGATCCATCGAAAGGAACGATTCGTGACCAGGTTCG 2163
     213 .....................................ProGlyPro 215

    2162 CGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCCCGATGTCGCCGCC 2113

     216 GluPro..............ArgThrArgLeuGlnProAlaThrProAr 227

    2112 TCGGCCG...ACTTCTACACCGGCCTGTTCGGCTGGACCGCCACCGTGGT 2066
     227 gArgSerGlyAlaAlaAspProAlaAspProValGlyHisPro......A 242

    2065 CAGCGACCCGGGCGCCGGGGGATACACGACGTTCAGCTCCGACGGGAAGC 2016
     242 laAlaProArgAlaProGlyPro.......................... 249

    2015 TGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCCCTACCACCGGCCG 1966
     250 ............GluProArgThrArgLeuGlnProAlaThrProArgAr 262

    1965 TACGGGCCCGGGAACGACCAGCACGGCATGC...CGGCCATCTGGACCGT 1919
     262 gSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProA 279

    1918 GTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGGGTCGAGACGG 1869
     279 rgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArg 295

    1868 CGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGGCCTCGGCCGG 1819
     296 ArgSerGlyAlaAlaAsp..............ProAlaAspProValGl 307

    1818 ATGGCGGTCTTCGCCGACCCCGCCGGGGCCGCGTTCGCGGTCTGGCGCAA 1769
     307 yHisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuG 324

    1768 GGGAGTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGCTCGGTCGGCT 1719
     324 lnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProVal 340

    1718 GGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCGGGACTTCTACCCG 1669
     341 GlyHisProAlaAlaProArgAlaProGlyPro.................. 351

    1668 GCGACCCTCGGCCTGGCTCCGGCCGACACCGGACTGAAGGGCGTCACCGA 1619
     352 .GluProArgThrArgLeuGlnPro..............AlaThrProA 363

    1618 CCCGGTCTGGCACATCG..............GTGACACACCGGTCG... 1587
     363 rgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAla 379

    1586 ..............CCGGCACCCAGGAGCTGGGCGTCACCGGCGCGGTA 1552
     380 ProArgAlaProGlyProGluProArgThrArgLeuGlnProAla..... 394

    1551 CGGCCGCACTGGGCCGTGCTGTTCGCCGTGCACGACTGCGACGCGACGGT 1502
     395 .....................................ThrProArgArgS 399

    1501 CCGGCGCGCCGTTGAACTCGGCGGCTCCGTCGAGAACGAGCCCGCCGACA 1452
     399 erGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArg 415

    1451 CGCCCAGGGGGCGGCGGGCGGACCTGCTCGACCCGCACGGGGCCGGCTTC 1402
     416 AlaProGlyProGluProArg.........ThrArgLeuGlnProAla.. 428

    1401 TCGGTGGTCGAACTGCGGGAGGGGTACCCCGCGGCGGCGGGCGGTGCCTC 1352
     429 ........................ThrProArg.................. 431

    1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
     432 .....ArgSerGlyAlaAlaAspProAlaAspProVal............ 442

    1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
```

```
 443 ...GlyHisProAla.............AlaProArgAlaProGly..... 452

1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
                          |||:::   ||||||||||||      :
 453 ......................ProGluProArgThrArgLeu......G 460

1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGC.................. 1170
     ::       ||||||     |||||||:::
 460 lnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProVal 476

1169 .........CGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCGCCG 1129
                ::::|||         ||||||:::||||||   |||
 477 GlyHisProAlaAlaProArgAlaProGlyProGluProArgThrArgLe 493

1128 CCTGGCGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTGCGT 1079
           ::      |||              ||||||||||||||:
 493 uGlnProAlaThrProArg..............ArgSerGlyAlaAlaA 505

1078 CCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCGCCG 1029
     ::                          :::|||   |||||||:::|||
 505 spProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyPro 521

1028 CATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGGATC 979
     :::||||||||||   :::||||||||||         |||:::   :::::
 522 GluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAlaAl 538

 978 GCATCGCCGCGAT......CGCCGGCCGGCTGCTCACCGAACT....... 942
     :         |||         :::|||||||||   |||:::
 538 aAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyP 555

 941 ........CGCCGACGCCTCCGGCCGGTC.....GGGCAAACCGGCCGA 906
             |||   ||||||:::|||:::         ::::::|||
 555 roGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAla 571

 905 GCTGATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGC 856
     ||||||               |||         |||||||||      ||
 572 AlaAsp..............ProAlaAspProValGlyHisProAlaAl 583

 855 TGCTCGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGC 806
     |   |||                  :::   ||||||:::   |||   |
 583 aProArg..................AlaProGlyProGluProArgThrA 594

 805 GTTCTCAAGGCACTCGGCCTCGGCGG..............CCCGCAGAG 771
     ||   |||   :::   |||||||||               |||||:::
 594 rgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAsp 610

 770 CGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCCGGACACCTCGG 721
           |||            |||:::      ||||||
 611 ProValGly.........HisProAlaAlaProArgAla........... 620

 720 CCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCCGGCGGAACGAC 671
      ||||||   |||   ||||||:::|||   |||:::||||||
 621 .ProGly...ProGluProArgThrArg...LeuGlnProAla....... 632

 670 ACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAGGCCGAGTTCGG 621
                    ||||||                          |||
 633 ...........ThrProArg..........................Arg 636

 620 CTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGGGCTCATCTTCG 571
        |||         ||||||            |||   |||      ||
 637 SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProAr 653

 570 CCGGCCA.........CGACACCACCGGCTCCTTCCTGGGCTTCCTGCTC 530
     |   |||           :::   ||||||   |||:::   |||   |
 653 gAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgA 670

 529 GCGGAGGTCCTGGCGGGCCG.........................CCT 507
     ||:::|||   :::   |||                         |||
 670 rgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaPro 686

 506 CGC......GGCGGATGCCGACGAGGACGCCGTCTCCCGGTT......CG 469
     |||      |||         |||   |||      |||:::      ||
 687 ArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProAr 703

 468 TGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTACACGTTGTGGAGG 419
     |   :::|||||||||   |||         :::   |||   :::
 703 gArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaP 720

 418 TTCGCTGCCACGGAGGTGACCATCGGCGGCGT......CCGGCTGCCCCG 375
       |||:::   |||   :::   |||   |||      ||||||:::|||
 720 roArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrPro 736

 374 CGGAGCGCCGGTGCTGGT.................GGACATCG...... 349
     |||   :::||||||:::                  ||||||
 737 ArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAl 753

 348 ....AGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTT 303
         |||||||||   |||                  |||||||||
 753 aProArgAlaProGlyPro..............GluProArgThrArgL 765

 302 CCACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGC 253
           ::::::||||||   :::|||:::   :::
 765 euGlnProAlaThrProArgArgSerGlyAlaAlaAsp............ 777

 252 ACTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATC 203
           |||:::      :::   |||:::      ||||||
 778 ...ProAlaAspProValGlyHisProAlaAlaProArgAla........ 790

 202 GGCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGA 153
                       |||   |||:::
 791 ......................ProGlyProGlu................ 794

 152 CGAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAAC 103
                            ||||||   |||   :::|||
 795 ..........................ProArgThrArgLeuGlnProAla 802

 102 TGCCCG......TCTGGCTGCGCTGATCCGCCGGCCGCGACCGGCACGGG 59
        |||      ||||||:::||||||||||      |||      :::||
 803 ThrProArgArgSerGlyAlaAlaAspPro......AlaAspProValGl 817

  58 ACCGCCCACCGCGCACCGCGCGGTGGGCGGTCCCGTGCCGTGTACCCGG 10
     |   |||:::|||   ||||||      ||||||   |||   ||||||
 817 yHisProAlaAlaProArgAla...ProGlyProGluProArgThrArg 832
```

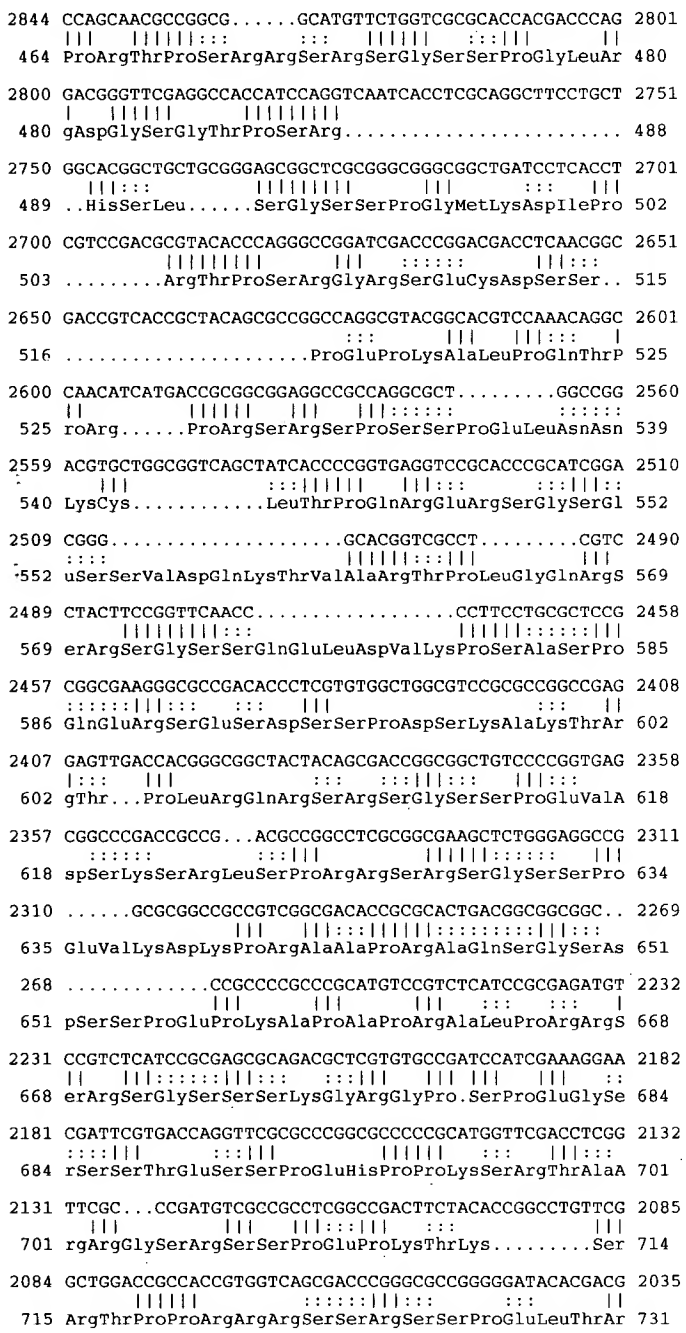
seq_name: pir2:T02345

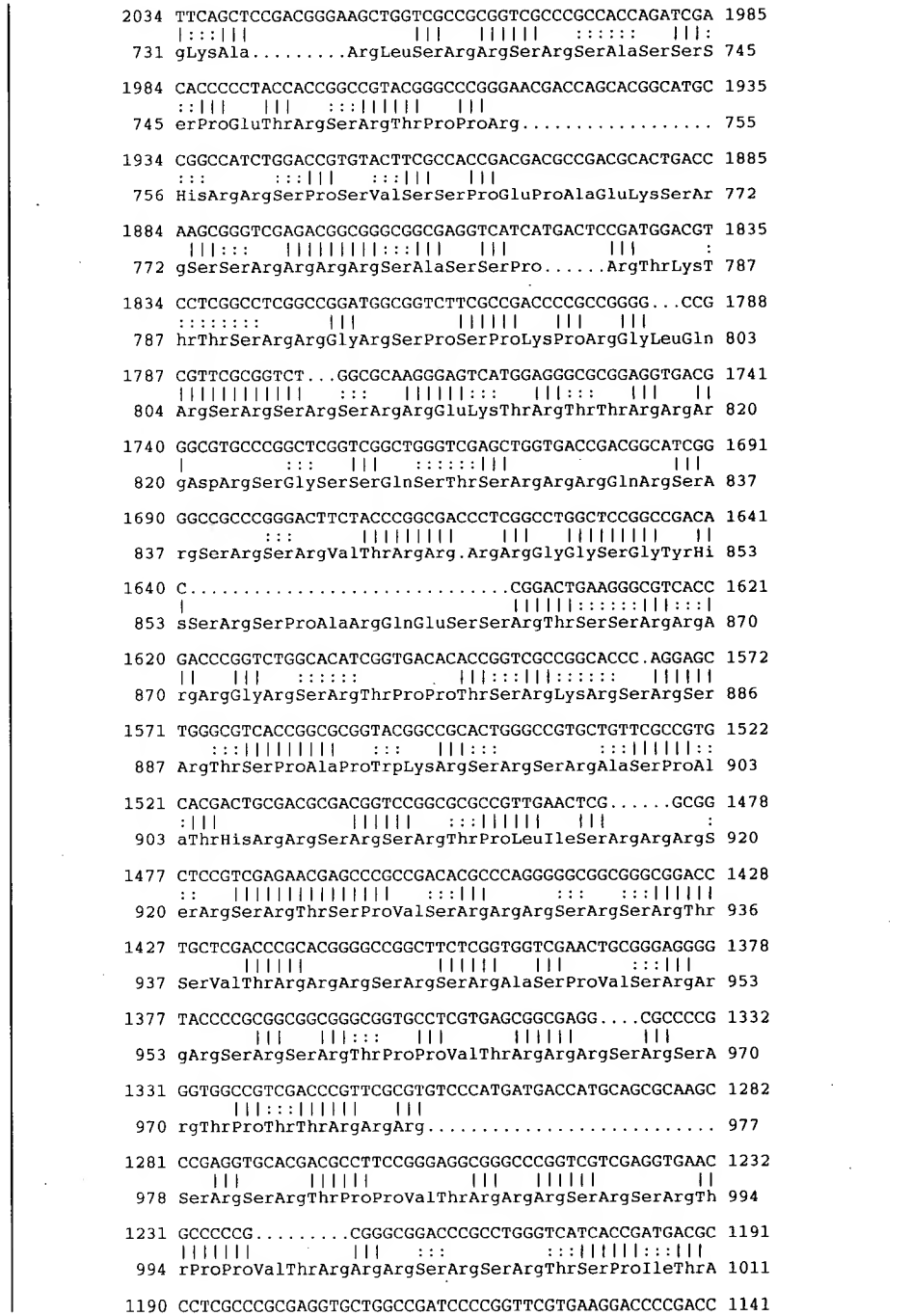
seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C;Accession: T02345
R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A;Description: Sequencing of human chromosome 16p13.3.
A;Reference number: Z14664
A;Accession: T02345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1791 <RIC>
A;Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C;Genetics:
A;Map position: 16
A;Introns: 1610/2; 1706/2
A;Note: KIAA0324

alignment_scores:
Quality: 486.50 Length: 1108
Ratio: 0.977 Gaps: 55
Percent Similarity: 44.946 Percent Identity: 27.978

alignment_block:
US-09-673-254-1/rev x T02345 ..

Align seg 1/1 to: T02345 from: 1 to: 1791


2844 CCAGCAACGCCGGCG......GCATGTTCTGGTCGCGCACCACGACCCAG 2801
464 ProArgThrProSerArgArgSerArgSerGlySerSerProGlyLeuAr 480
2800 GACGGGTTCGAGGCCACCATCCAGGTCAATCACCTCGCAGGCTTCCTGCT 2751
480 gAspGlySerGlyThrProSerArg..... 488
2750 GGCACGGCTGCTGCGGGAGCGGCTCGCGGGCGGGCGGCTGATCCTCACCT 2701
489 ..HisSerLeu......SerGlySerSerProGlyMetLysAspIlePro 502
2700 CGTCCGACGCGTACACCCAGGGCCGGATCGACCCGGACGACCTCAACGGC 2651
503 .....ArgThrProSerArgGlyArgSerGluCysAspSerSer.. 515
2650 GACCGTCACCGCTACAGCGCCGGCCAGGCGTACGGCACGTCCAAACAGGC 2601
516 .....ProGluProLysAlaLeuProGlnThrP 525
2600 CAACATCATGACCGCGGCGGAGGCCGCCAGGCGCT.....GGCCGG 2560
525 roArg......ProArgSerArgSerProSerSerProGluLeuAsnAsn 539
2559 ACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTCCGCACCCGCATCGGA 2510
540 LysCys.....LeuThrProGlnArgGluArgSerGlySerGl 552
2509 CGGG.....GCACGGTCGCCT.....CGTC 2490
552 uSerSerValAspGlnLysThrValAlaArgThrProLeuGlyGlnArgS 569
2489 CTACTTCCGGTTCAACC.....CCTTCCTGCGCTCCG 2458
569 erArgSerGlySerSerGlnGluLeuAspValLysProSerAlaSerPro 585
2457 CGGCGAAGGGCGCCGACACCCTCGTGTGGCTGGCGTCCGCGCCGGCCGAG 2408
586 GlnGluArgSerGluSerAspSerSerProAspSerLysAlaLysThrAr 602
2407 GAGTTGACCACGGGCGGCTACTACAGCGACCGGCGGCTGTCCCCGGTGAG 2358
602 gThr...ProLeuArgGlnArgSerArgSerGlySerSerProGluValA 618
2357 CGGCCCGACCGCCG...ACGCCGGCCTCGCGGCGAAGCTCTGGGAGGCCG 2311
618 spSerLysSerArgLeuSerProArgArgSerArgSerGlySerSerPro 634
2310 ......GCGCGGCCGCCGTCGGCGACACCGCGCACTGACGGCGGCGGC.. 2269
635 GluValLysAspLysProArgAlaAlaProArgAlaGlnSerGlySerAs 651
268 .....CCGCCCCGCCCGCATGTCCGTCTCATCCGCGAGATGT 2232
651 pSerSerProGluProLysAlaProAlaProArgAlaLeuProArgArgS 668
2231 CCGTCTCATCCGCGAGCGCAGACGCTCGTGTGCCGATCCATCGAAAGGAA 2182
668 erArgSerGlySerSerSerLysGlyArgGlyPro.SerProGluGlySe 684
2181 CGATTCGTGACCAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGG 2132
684 rSerSerThrGluSerSerProGluHisProProLysSerArgThrAlaA 701
2131 TTCGC...CCGATGTCGCCGCCTCGGCCGACTTCTACACCGGCCTGTTCG 2085
701 rgArgGlySerArgSerSerProGluProLysThrLys.....Ser 714
2084 GCTGGACCGCCACCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACG 2035
715 ArgThrProProArgArgArgSerSerArgSerSerProGluLeuThrAr 731


2034 TTCAGCTCCGACGGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGA 1985
731 gLysAla.....ArgLeuSerArgArgSerArgSerAlaSerSerS 745
1984 CACCCCCTACCACCGGCCGTACGGGCCCGGGAACGACCAGCACGGCATGC 1935
745 erProGluThrArgSerArgThrProProArg..... 755
1934 CGGCCATCTGGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACC 1885
756 HisArgArgSerProSerValSerSerProGluProAlaGluLysSerAr 772
1884 AAGCGGGTCGAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGT 1835
772 gSerSerArgArgArgArgSerAlaSerSerPro.....ArgThrLysT 787
1834 CCTCGGCCTCGGCCGGATGGCGGTCTTCGCCGACCCCGCCGGGG...CCG 1788
787 hrThrSerArgArgGlyArgSerProSerProLysProArgGlyLeuGln 803
1787 CGTTCGCGGTCT...GGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACG 1741
804 ArgSerArgSerArgSerArgArgGluLysThrArgThrThrArgArgAr 820
1740 GGCGTGCCCGGCTCGGTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGG 1691
820 gAspArgSerGlySerSerGlnSerThrSerArgArgArgGlnArgSerA 837
1690 GGCCGCCCGGGACTTCTACCCGGCGACCCTCGGCCTGGCTCCGGCCGACA 1641
837 rgSerArgSerArgValThrArgArg.ArgArgGlyGlySerGlyTyrHi 853
1640 C.....CGGACTGAAGGGCGTCACC 1621
853 sSerArgSerProAlaArgGlnGluSerSerArgThrSerSerArgArgA 870
1620 GACCCGGTCTGGCACATCGGTGACACACCGGTCGCCGGCACCC.AGGAGC 1572
870 rgArgGlyArgSerArgThrProProThrSerArgLysArgSerArgSer 886
1571 TGGGCGTCACCGGCGCGGTACGGCCGCACTGGGCCGTGCTGTTCGCCGTG 1522
887 ArgThrSerProAlaProTrpLysArgSerArgSerArgAlaSerProAl 903
1521 CACGACTGCGACGCGACGGTCCGGCGCGCCGTTGAACTCG.....GCGG 1478
903 aThrHisArgArgSerArgSerArgThrProLeuIleSerArgArgArgS 920
1477 CTCCGTCGAGAACGAGCCCGCCGACACGCCCAGGGGGCGGCGGGCGGACC 1428
920 erArgSerArgThrSerProValSerArgArgArgSerArgSerArgThr 936
1427 TGCTCGACCCGCACGGGGCCGGCTTCTCGGTGGTCGAACTGCGGGAGGGG 1378
937 SerValThrArgArgArgSerArgSerArgAlaSerProValSerArgAr 953
1377 TACCCCGCGGCGGCGGGCGGTGCCTCGTGAGCGGCGAGG....CGCCCCG 1332
953 gArgSerArgSerArgThrProProValThrArgArgArgSerArgSerA 970
1331 GGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGC 1282
970 rgThrProThrThrArgArgArg..... 977
1281 CCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAAC 1232
978 SerArgSerArgThrProProValThrArgArgArgSerArgSerArgTh 994
1231 GCCCCCG.....CGGGCGGACCCGCCTGGGTCATCACCGATGACGC 1191
994 rProProValThrArgArgArgSerArgSerArgThrSerProIleThrA 1011
1190 CCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACC 1141

```
          :::||| :::      |||||||||   :::      |||
1011 rgArgArgSerArg.SerArgThrSerProValThrArg...ArgArgSe 1026

1140 TCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCGACATCCCCGTT 1091
      |||   |||                  |||||||||          ||||
1026 rArgSerArgThrSerProValThrArgArgArgSer.........ArgS 1040

1090 CCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCG 1041
     ||   ::::::   |||      :::|||         |||   ||||||
1040 erArgThrSerProValThrArgArgArgSerArgSerArgThrProPro 1056

1040 GCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCG 991
     |||         :::   |||:::   :::        |||      ||
1057 AlaIleArgArgArgSerArgSerArgThrProLeuLeuProArgLysAr 1073

 990 AGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTC 941
     |:::              |||:::|||   |||   |||   |||   |
1073 gSer..............ArgSerArgSerProLeuAlaIleArgArgA 1085

 940 GCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTT 891
     ||   |||   |||   :::   :::|||:::|||:::   |||   |||
1085 rgSerArgSerArg...ThrProArgThrAlaArgGlyLysArgSerLeu 1100

 890 CGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGG 841
                |||||||:::   :::      |||:::      :::::
1101 ThrArgSerProProAlaIleArgArgArgSerAlaSerGlySerSerSe 1117

 840 TCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTC 791
     :   ||||||  |||      |||   |||   |||      :::|||||
1117 rAspArgSer..Arg...SerAlaThrProProAlaThrArgAsnHisSe 1132

 790 GGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGG 741
     |:::|||                                |||||   |||
1132 rGlySer.............................ArgThrProProV 1139

 740 GGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCG... 694
       |||   ::::::   |||       :::|||   |||:::|||
1139 alAlaLeuAsnSerSerArg......MetSerCysPheSerArgProSer 1153

 693 .......................TGCACTCAGCCCGGCGGAACGACACC 668
                           |||   :::||||||         ||
1154 MetSerProThrProLeuAspArgCysArgSerProGlyMetLeuGluPr 1170

 667 CCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTC 618
     |                      |||:::|||   |||   |||:::
1170 o..................LeuGlySerSerArgThrProMetSerValL 1181

 617 GGTCTCCGACGACCAGCTCGTCTACATGATCACCGGGCTCATCTTCGCCG 568
       :::   :::   :::|||        :::||||||        |||
1181 euGlnGlnAlaGlyGlySerMetMetAspGlyProGlyProArgIlePro 1197

 568 ..................................................  568

1198 AspHisGlnArgThrSerValProGluAsnHisAlaGlnSerArgIleAl 1214

 567 .......................GCCACGACACCACCGGCTCCTTCCT 543
                           :::|||:::   |||   ||||||
1214 aLeuAlaLeuThrAlaIleSerLeuGlyThrAlaArgProProProSerM 1231

 542 GGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCG 493
       ::::::::::::                 |||:::|||
1231 etSerAlaAlaGly..............LeuAlaAlaArgMetSerGln 1242

 492 ACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCG 443
     :::   :::|||   |||                        :::   ||
1243 ValProAlaProValPro.................LeuMetSerLeuAr 1253

 442 CCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGG 393
     |   :::|||         :::::::::|||      |||   |||:::|
1253 gThrAlaPro.........AlaAlaAsnLeuAlaSerArgIleProAlaA 1267

 392 CGGCGTCCGGCTGC.....................CCCGCGGAGCGC 367
     ||::::::::::::                     ||||||
```

```
1267 laSerAlaAlaAlaMetAsnLeuAlaSerAlaArgThrProAlaIlePro 1283

 366 CGGTGCTGGTGGACATCGAGGGCACCAACACCGACGGCCGCC........ 325
       :::         :::   :::   ||||||:::||||||
1284 ThrAlaValAsnLeuAlaAspSerArgThrProAlaAlaAlaAlaAlaMe 1300

 324 ....ATCACGACGCCCCGCACG.......................... 307
         ::::::::::|||||||||
1300 tAsnLeuAlaSerProArgThrAlaValAlaProSerAlaValAsnLeuA 1317

 306 .....CCTTCCACCCGGACCGTCCCT............CGTGGCGGCGGC 274
          |||   |||   |||:::|||               |||:::
1317 laAspProArgThrProThrAlaProAlaValAsnLeuAlaGlyAlaArg 1333

 273 TCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGCTCGCCCAG 224
     :::|||:::|||   :::       :::   :::||||||::::::|||
1334 ThrProAlaAlaLeuAlaAlaLeuSerLeuThrGlySerGlyThrProPr 1350

 223 CTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTCCCCGAGGC 174
                          :::|||         :::::|||      |
1350 oThr..............AlaAlaAsnTyrProSerSerSerArgThrP 1362

 173 CCGACTGGCCGTGCCGTACGACGAGTTGCGGT.................. 142
     ||:::   |||:::   ::::::      |||
1362 roGlnAlaProAlaSerAlaAsnLeuValGlyProArgSerAlaHisAla 1378

 142 .................................................. 142

1379 ThrAlaProValAsnIleAlaGlySerArgThrAlaAlaAlaLeuAlaPr 1395

 141 .............GGTGCCGGAAGGGGGCCCAGACGGCGCGGC....... 112
                      :::|||      :::|||         |||
1395 oAlaSerLeuThrSerAlaArgMetAlaProAlaLeuSerGlyAlaAsnL 1412

 111 .....TCACCGAACTGCCCG.....................TCTGGCTGC 88
          ||||||:::   |||                      ||||||
1412 euThrSerProArgValProLeuSerAlaTyrGluArgValSerGlyArg 1428

  87 GCTGATCCGCCG.........GCCGCGACCGGCACGGGACCGCCC..... 52
     ::::::||||||            |||   :::      ||||||
1429 ThrSerProProLeuLeuAspArgAlaArgSerArgThrProProSerAl 1445

  51 ...................ACCGCGCACCGCGCGGTGGGCGGTCCCGTGC 21
                          |||:::::||||||         |||   |
1445 aProSerGlnSerArgMetThrSerGluArgAla.........ProSerP 1459

  20 CGTGTACCCGGTGCGGA 4
     ||   :::|||   |||
1459 roSerSerArgMetGly 1464
```

seq_name: pir2:T02345

seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C;Accession: T02345
R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A;Description: Sequencing of human chromosome 16p13.3.
A;Reference number: Z14664
A;Accession: T02345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1791 <RIC>
A;Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C;Genetics:
A;Map position: 16
A;Introns: 1610/2; 1706/2
A;Note: KIAA0324

```
alignment_scores:
            Quality:  479.00              Length:   1056
              Ratio:  1.028                Gaps:     56
 Percent Similarity:  44.129   Percent Identity: 29.261

alignment_block:
 US-09-673-254-1 x T02345   ..

 Align seg 1/1  to: T02345  from: 1  to: 1791

     133 TCCGGCACCACCGCAACTCGTCGTACGGCACGGCCAGTC.........GG 173
         :::|||   ::::::::::::       :::   ||||||
     426 AlaGlyMetSerSerAsnGlnSerIleSerSerProValLeuAspAlaVa 442

     174 GCCTCGGGGAACCTGCTGCGCAGTACGCCGATCATCGTGCGCGACTCCAG 223
          ||||||   |||             |||   ||||||:::::::::|||
     442 lProArg...ThrProSerArgGluArgSerSerSerAlaSerSerProG 458

     224 CTGGGCGA...GCTGCTCCCCGATGCAGTAGTGCGGCCCGTCGCCGAAGG 270
             :::   :::   ||||||      |||          |||
     458 luMetLysAspGlyLeuProArgThrProSerArgArgSerArgSerGly 474

     271 TGAGCCGCCGCCACGAGGGACGGTCCGGGT..................... 300
           :::   :::   |||||||||   |||
     475 SerSerProGlyLeuArgAspGlySerGlyThrProSerArgHisSerLe 491

     301 ....GGAAGGCGTGCGGGGCGTCGTGATGGCGGCCGT............. 333
             |||         |||   :::|||      |||
     491 uSerGlySerSerProGlyMetLysAspIleProArgThrProSerArgG 508

     334 ..CGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCCGCGGGGCA 381
         |||      |||         |||         ||||||      :::
     508 lyArgSerGluCysAspSerSerProGluProLysAlaLeuProGlnThr 524

     382 GCCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGAACCTCCACAACGTG 431
               |||      ||||||         :::   :::::   ||
     525 ProArgProArgSerArgSerProSerSerProGluLeuAsnAsnLysCy 541

     432 TAGGGCACCG............GCGGGTGGTAGCGCAGCGCCTCCTCCAC 469
         |   :::|||           :::|||:::::::::::::        :
     541 sLeuThrProGlnArgGluArgSerGlySerGluSerSerValAspGlnL 558

     470 GAACCGGGAGACGGCGTCCTCGTCGGCATCCGCCGCGAGGCGGC...... 513
         ::|||      |||   |||      :::      |||:::|||
     558 ysThrValAlaArgThrProLeuGlyGlnArgSerArgSerGlySerSer 574

     513 ................................................. 513

     575 GlnGluLeuAspValLysProSerAlaSerProGlnGluArgSerGluSe 591

     514 ..........CCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGGAGCC 553
                    |||   :::      ::::::   :::|||   |||:::|
     591 rAspSerSerProAspSerLysAlaLysThrArgThrProLeuArgGlnA 608

     554 GGTGGTGTCGTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACGAGCT 603
         ||       |||         :::      :::   |||   :::
     608 rgSerArgSerGlySerSerProGluValAspSerLysSerArgLeuSer 624

     604 GGTCGTCGGAGACCGAGCCGAACTCGGCCTGCG.........CGCGCTCG 644
           ||||||      |||   :::   |||             :::   ||
     625 ProArgArgSerArgSerGlySerSerProGluValLysAspLysProAr 641

     645 TACAGCACGCGGGTCATGGTCGGGGTGTCGTTCCGCCGGGCTGAGTGCAC 694
         |:::|||      :::   ||||||      |||
     641 gAlaAlaProArgAlaGlnSerGlySerAspSer.............. 652

     695 GGCTTCGAGGAGCAGGCTCTCCAGGGCCGAGGTGTCCGGCACGCCCCCGG 744
                         ||||||   |||:::::::||||||   ||||||
     653 ................SerProGluProLysAlaProAla...ProArg 662

     745 CAGGGTCCGTGCCGTCACCCCCGCCGCTCTGCGGGCCGCCGAGGCCGAGT 794
               |||   |||:::   |||   :::::::::   :::||||||
     663 AlaLeuPro...ArgArgSerArgSerGlySerSerSerLysGlyArgGl 678

     795 GCCTTGAGAACGCTGACGGCCTCGCGGG..............CCATCGC 829
          |||               |||:::|||                 ||||
     678 yPro............SerProGluGlySerSerSerThrGluSerSerP 691

     830 CGGATCGGTGACCGGCACACCGAGCAGCTCGCAGATGACCAACAGCGGGA 879
         ||:::::   |||       |||:::|||||||||
     691 roGluHisProProLysSerArgThrAlaArgArg............... 702

     880 AGTGGTACGCGAAGCCGCCGATCAGCTCGGCCGGTTTGCCCGACCGGCCG 929
                 ||||||   |||   :::|||
     703 .........GlySerArgSerSerProGluPro................. 710

     930 GAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCCGGCGATCGCGGCGATGCG 979
             :::   :::   |||            |||||||||   |||   :
     711 ....LysThrLysSerArgThrProProArgArgArgSerSerArgSerS 726

     980 ATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGAACGCAGGTGCGTGGATGC 1029
         ::|||::: |||   :::   ||||||:::::||||||      :::||
     726 erProGlu.LeuThrArgLysAlaArgLeuSerArgArgSerArgSerAl 742

    1030 GGCGCAGGCGCCGGTGGGCCTCGCCGTCCACGGCGATGAGCGTGAACGGA 1079
         |::::::::::|||                  :::   |||:::||||
     742 aSerSerSerPro.....................GluThrArgSerArgT 752

    1080 CGCAGCTCCGGAACGGGGATGTCGAGACCGTCGT....CCACCCCCCGCC 1125
         ||      |||:::            :::|||      |||   ||||||
     752 hrProProArgHisArgArgSerProSerValSerSerProGluProAla 768

    1126 ..................AGGCGGCGGGGGCGAGGTCGGGGTCCT..... 1152
                           |||||||||      :::   :::|||
     769 GluLysSerArgSerSerArgArgArgArgSerAlaSerSerProArgTh 785

    1153 ..........TCACGAACCGGGGATCGGCCAGCACCTCGCGGGCGAGGGC 1192
                   ||||||   |||      |||:::|||:::   ||||||
     785 rLysThrThrSerArgArgGlyArgSerProSerProLysProArgGlyL 802

    1193 GTCATCGGTGATGACCCAGGCGGGTCCGCCCGCGGGGGCGTTCACCTCGA 1242
           :::|||         |||      |||   :::   |||:::   |||
     802 euGlnArgSerArgSerArgSerArgArgGluLysThrArgThrThrArg 818

    1243 CGACCGGGCCCGCCTCCCGGAAGGCGTCGTGCACCTCGGGCTTGCGCTGC 1292
         |||                  |||   |||
     819 Arg..................ArgAspArg..................... 822

    1293 ATGGTCATCATGGGACACGCGAACGGGTCGACGGCCACCCGGGGCGCCTC 1342
             |||:::   :::      |||:::||||||         :::
     823 ....SerGlySerSerGlnSerThrSerArgArgArgGlnArgSerArgS 838

    1343 GCCGCTCACGAGGCACCGCCCGCCGCCGCGGGG.........TACCCCTC 1383
           |||||||||   |||   |||||||||  |||         |||   ||
     838 erArgSerArgValThrArgArgArgArg.GlyGlySerGlyTyrHisSe 854

    1384 CCGC..................AGTTCGACCACCGAGAAGCCGGCCCCGT 1415
         ||||                  ||||||   |||::: ||||||   |||
     854 rArgSerProAlaArgGlnGluSerSerArgThrSer.SerArgArgArg 870

    1416 GCGGGTCGAGCAGGTCCGCCCGCCGCCCCCTGGGCGTGTCGGCGGGCTCG 1465
           ||||||:::      |||      |||   :::   :::|||:::||
     871 ArgGlyArgSerArgThrPro......ProThrSerArgLysArgSerAr 885

    1466 TTCTCGACGGAGC.............CGCCGAGTTCAACGGCGCGCCGGAC 1503
         |||||||   |||            :::|||   :::   ||||||:::|
     885 gSerArgThrSerProAlaProTrpLysArgSerArgSerArgAlaSerP 902

    1504 CGTCG......CGTCGCAGTCGTGCACGGCGAACAGCACGGCCCAGTGCG 1547
         ||:::      ||||||||||||:::|||           |||
     902 roAlaThrHisArgArgSerArgSerArg.........ThrProLeuIle 915

    1548 GCCGTACCGCGC...CGGTGACGCCCAGCTCCTGGGTGCCGGCGACCGGT 1594
         :::      |||   |||   |||   :::|||      |||
     916 SerArgArgArgSerArgSerArgThrSerProValSerArg........ 929
```

```
1595 GTGTCACCGATGTGCCAGACCGGGTCGGTGACGCCCTTCAGTCCGGTGTC 1644
                  |||   :::|||   |||   ||||||      |
 930 ..............ArgArgSerArgSerArgThrSerValThrArgA 941

1645 GGCCGGAGCCAGGCCGAGGGTCGCCG.........GGTAGAAGTCCCGGG 1685
     ||   :::   :::|||:::||||||           ||||||   :::
 941 rgArgSerArgSerArgAlaSerProValSerArgArgArgSerArgSer 957

1686 CGG...CCCCGATGCCGTCGGTCACCAGCTCGACCCAGCCGACCGAGCCG 1732
     |||   |||        |||      :::|||         |||:::
 958 ArgThrProProValThrArgArgArgSerArgSerArgThrProThrTh 974

1733 GGCACGCCCGTCACCTCCGCGCCCTCCATGACTCCCTTGCGCCAGACCGC 1782
       |||   |||      |||   |||                   |||
 974 rArgArgArgSerArgSerArgThrPro.................ProV 985

1783 GAACGCGGCCCCGGCGGGGTCGGCGAAGACCGCCATCCGGCCGAGGCCGA 1832
       ||||||   |||         |||   ||||||      |||   |||
 985 alThrArgArgArgSerArgSerArgThrProProValThrArgArgArg 1001

1833 GGACGTCCATCGGAGTCATGATGACCT.........CGCCGCCCGCCGTC 1873
     :::|||        |||      |||         ||||||      ||
1002 SerArgSerArgThrSer......ProIleThrArgArgArgSerArgSe 1016

1874 TCGACCCGCTTGGTCAGTGCGTCGGCGTCGTCGGTGGCGAAGTACACGGT 1923
     ||||   :::      :::||||||||||   |||   |||::::::
1016 rArgThrSerProValThrArgArgArgSerArgSerArgThrSer.... 1031

1924 CCAGATGGCCGGCATGCCGTGCTGGTCGTTCCCGGGCCCGTACGGCCGGT 1973
          |||:::   |||      |||||||||:::||||||:::
1032 ........ProValThrArgArg...ArgSerArgSerArgThrSerPro 1044

1974 GGTAGGGGGTGTCGATCTGGTGGCGGGCGACCGCGGCGACCAGCTTCCCG 2023
               ||||||   :::      |||      ||||||   ||
1045 ValThrArgArgArgSerArgSerArgThrPro......ProAlaIleAr 1059

2024 T.........CGGAGCTGAACGTCGTGTATCCCCCGGCGCCCGGGTCGCT 2064
     |         ||||||   |||      :::||||||:::
1059 gArgArgSerArgSerArgThrProLeuLeuProArgLys.......... 1072

2065 GACCACGGTGGCGGTCCAGCCGAACAGGCCGGTGTAGAAGTCGGCCG... 2111
                              |||   |||||||||
1073 ............................ArgSerArgSerArgSerPro 1079

2112 .........AGGCGGCGACATCGGGCGAACCGAGGTCGAACCATGCGGGG 2152
             |||||||||   |||:::::      ||||||   |||||
1080 LeuAlaIleArgArgArgSerArgSerArgThrProArgThrAlaArgGl 1096

2153 GCGCCGGGCGCGAACCTGGTCA............CGAATCGTTCCTTTCG 2190
     |:::|||::   |||   |||             |||           :
1096 yLysArgSerLeuThrArgSerProProAlaIleArgArgArgSerAlaS 1113

2191 ATGGATCGGCACACGAGCGTCTGCGCTCGCGGATGAGACGGACATCTCGC 2240
     ::||||||:::   :::|||   ||||||            ::::::|||
1113 erGlySerSerSerAspArgSerArgSerAlaThrProProAlaThrArg 1129

2241 GGATGAGACGGACATGCGGGCGGGGCGGGCCGCCGCCGTCAGTGCGCGG. 2289
     :::   :::|||    |||:::     :::|||      |||:::|||
1130 AsnHisSerGly.SerArgThrProProValAlaLeuAsnSerSerArgM 1146

2290 .....TGTCGCCGACGG.........................CGGCCG 2307
          |||      |||                         |||
1146 etSerCysPheSerArgProSerMetSerProThrProLeuAspArgCys 1162

2308 CGCCGGCCTCCCAGAGCTTCGCCGCGAGGCCGGCGTCGGCGGTCGGGCCG 2357
     |||   |||         :::|||   |||            |||:::||
1163 ArgSerProGlyMetLeuGluProLeuGly............SerSerAr 1175

2358 CTCACCGGGGACAGCCGCCGGTCGCTGTAGTAGCCGCCCGTGGTCAACTC 2407
     |:::|||   ::::::     ::::::::::::       :::   |
1175 gThrProMetSerValLeuGlnGlnAlaGlyGlySerMetMetAspGlyP 1192

2408 CTCGGCCG..................................... 2415
```

```
     ||   |||
1192 roGlyProArgIleProAspHisGlnArgThrSerValProGluAsnHis 1208

2416 .....................GCGCGGACGCCAGCCACACGAGGGTGTCG 2444
                           |||   |||      :::   |||
1209 AlaGlnSerArgIleAlaLeuAlaLeuThrAlaIleSerLeuGlyThrAl 1225

2445 GCGCCCTTCGCCGCGGAGCGCAGGAAGGGGTTGAACCGGAAGTAGGACGA 2494
       ||||||   |||       ::::::   |||   ::::::       ::::
1225 aArgProProProSerMetSerAlaAlaGlyLeuAlaAlaArgMetSerG 1242

2495 GGCGACCGTGCCCCGTCCG....ATGCGGGTGCGGACCTCACCGGGGTGA 2540
     ::   |||:::||||||      |||   :::||||||:::|||
1242 lnValProAlaProValProLeuMetSerLeuArgThrAlaPro...... 1256

2541 TAGCTGACCGCCAGCACGTCCGGCCAGCGCCTGGCGGCCTCCGCCGCGGT 2590
           :::|||:::   :::::::::       ||||||||||||||::
1257 ......AlaAlaAsnLeuAlaSerArgIleProAlaAlaSerAlaAlaAl 1271

2591 CATGATGTTGGCCTGTTTGGACGTGCCGTACGCCTGGCCGGCGCTGTAGC 2640
     :|||   ||||||          :::|||
1271 aMetAsnLeuAlaSerAlaArgThrPro..................... 1280

2641 GGTGACGGTCGCCGTTGAGGTCGTCCGGGTCGATCCGGCCCTGGGTGTAC 2690
                                    :::|||         |||
1281 ............................AlaIleProThrAlaValAsn 1287

2691 GCGTCGGACGAGGTGAGGATCAGCCGCCCGCCCGCGAGCCGCTCCCGCAG 2740
        :::|||            ||||||   ||||||:::   :::
1288 LeuAlaAsp............SerArgThrProAlaAlaAlaAlaAlaMe 1300

2741 CAGCCGTGCCAGCAGGAAGCCTGCGAGGTGATTGACCTGGATGGTGGCCT 2790
       :::   ||||||   :::   |||                  ||||||
1300 tAsnLeuAlaSerProArgThrAla.................ValAla. 1310

2791 CGAACCCGTCCTGGGTCGTGGTGCGCGACCAGAACATGCCG......... 2831
          ||||||   |||   :::   |||   :::   |||
1311 .....ProSerAlaValAsnLeuAlaAspProArgThrProThrAlaPro 1325

2832 ........................CCGGCGTTGCTGGCCATGACATC 2854
                             ||||||   ||||||      ||
1326 AlaValAsnLeuAlaGlyAlaArgThrProAlaAlaLeuAlaAlaLeuSe 1342

2855 GATGCGCGGG 2864
     |:::   |||
1342 rLeuThrGly 1345
```

seq_name: pir1:B45344

seq_documentation_block:
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C;Species: suid herpesvirus 1
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: B45344
R;Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzer, M.
Virology 179, 365-377, 1990
A;Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented
A;Reference number: A45344; MUID:91021039
A;Accession: B45344
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-1733 <VLC>
A;Cross-references: GB:M34651; NID:g334070; PIDN:AAA47471.1; PID:g334072
C;Superfamily: pseudorabies virus 1 nuclear antigen

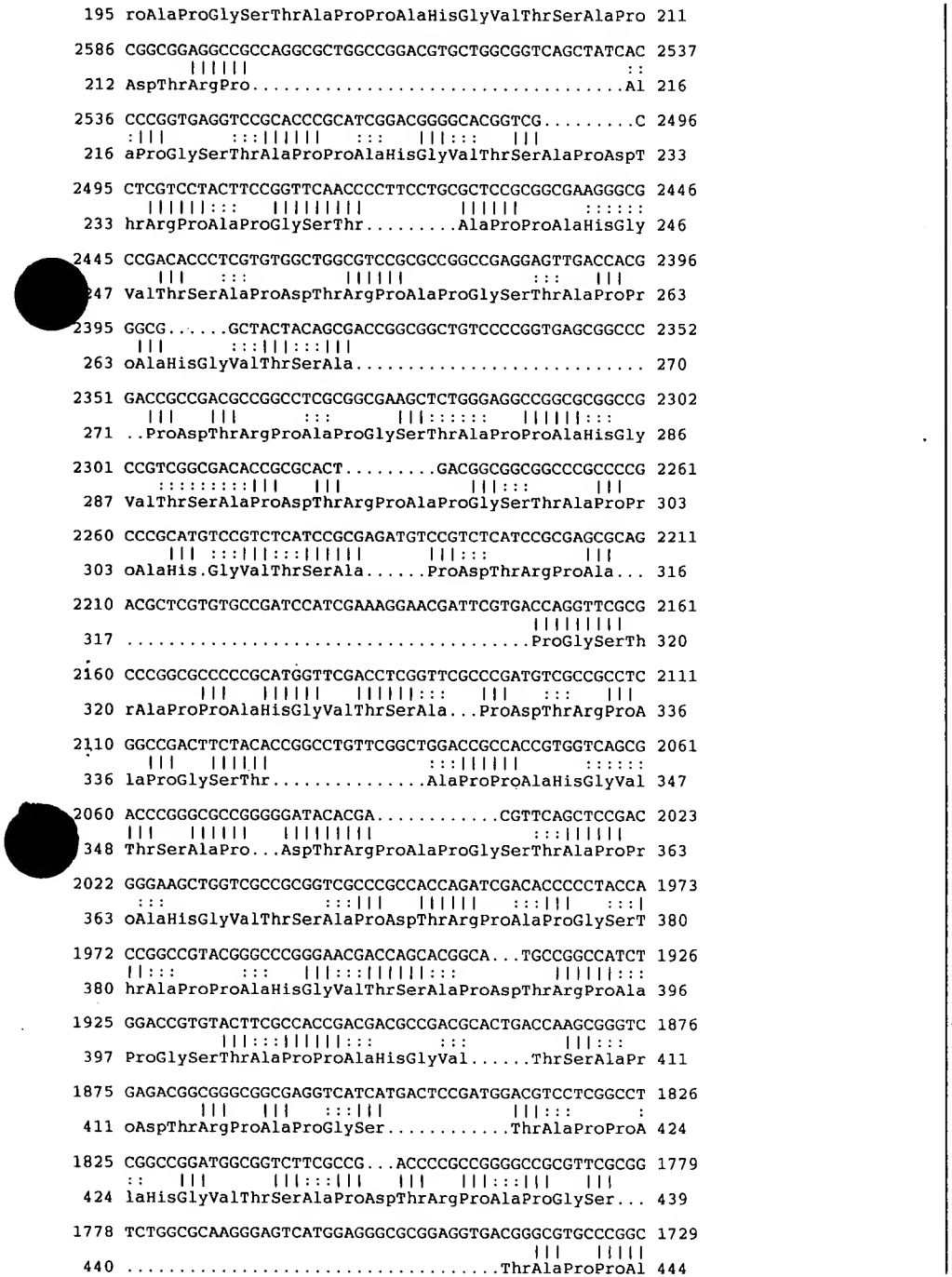
alignment_scores:
Quality: 469.00 Length: 1226
Ratio: 0.925 Gaps: 72
Percent Similarity: 41.354 Percent Identity: 29.364

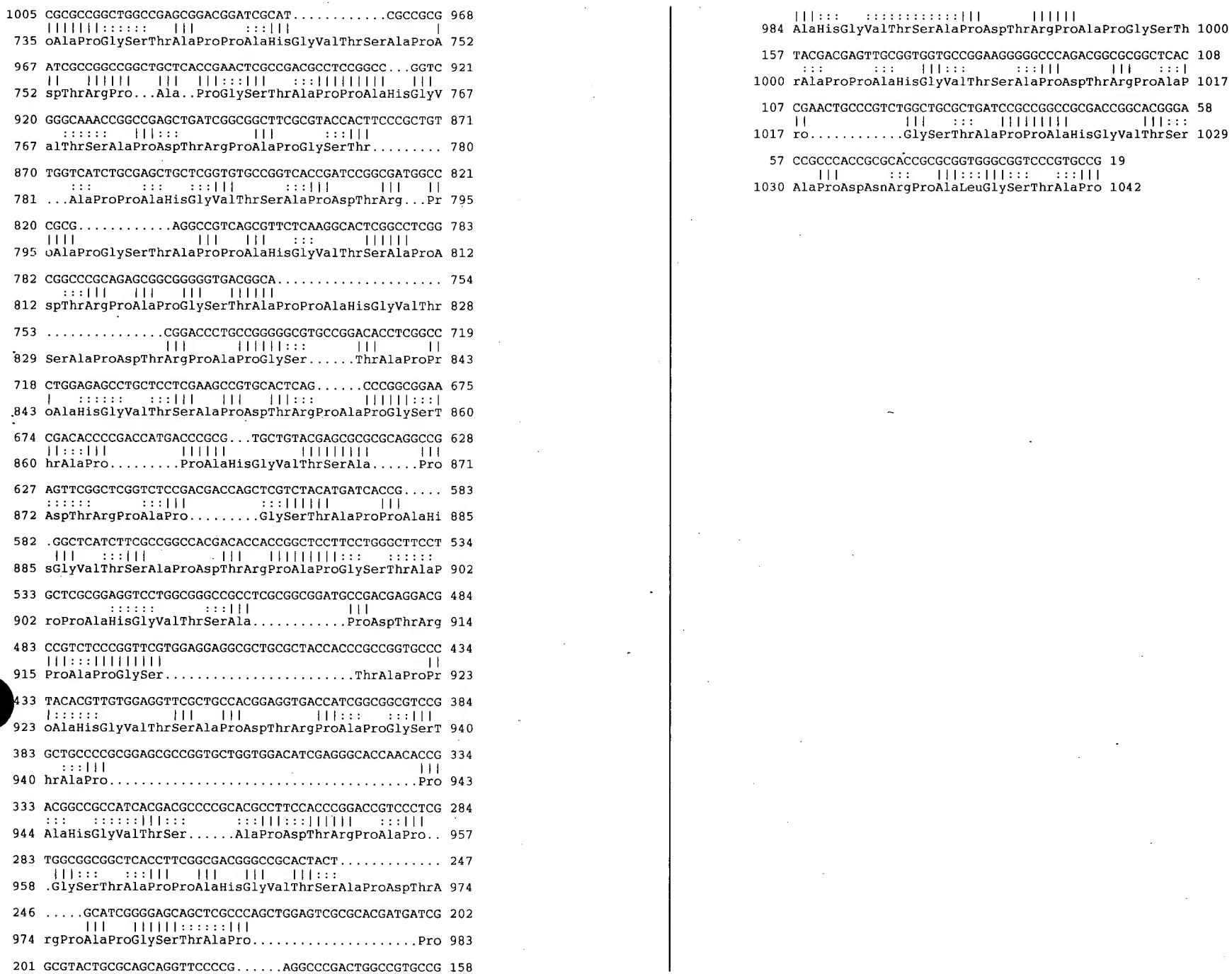
alignment_block:
US-09-673-254-1 x B45344 ..

```
Align seg 1/1  to: B45344  from: 1  to: 1733

      2 GATCCGCACCGGGTA...CACGGCACGGGACCGCCCACCGCGCGGTGCGC 48
        |||      ||||||   ||||||                 |||:::   ||
    631 AspGlyAlaArgValGlnHisGly.............ArgAlaGluLeuAl 643

     49 GGTGGGCGGTCCCGTGCCGGTCGCGGCCGGCGGATCAGCGCAGCCAGACG 98
        |:::   ||||||:::      |||   ||||||   :::|||   |||
    643 aAlaProGlyProAlaHisGlyAlaLeuGlyGlyArgValGlnAlaAsp. 659

     99 GGCAGTTCGGTGAGCCGCGCCGTCTGGGCCCCCTTCCGGCACCACCGCAA 148
                                           |||      :::
    660 ....................................AlaAspValAsp 663

    149 CTCGTCGTACGGCACGGCCAGTCGGGCCTCGGG..................... 181
        :::||||||   ||||||:::::      |||
    664 ValValValProHisGlyArgAlaValArgGlyProValLeuAspGlyVa 680

    182 ..........GAACCTGCTGCGCAGTACGCCGATCATCGTGCGCGACTCC 221
                 |||||||||   :::   |||:::   ||||||:::
    680 lGlnHisAspGluProAlaProArgArgAlaGluProArgAlaGlu.... 695

    222 AGCTGGGCGAGCTGCTCCCCGATGCAGTAGTGCGGCCCGTCGCCGAAGGT 271
                                 |||:::::   ::::::|||:::
    696 ........................ValLeuHisGlyAlaGlyGluAla 703

    272 GAGCCGCCGCCACGA..................................... 286
        |||   |||   |||
    704 GluValProArgArgGluGlnGlnHisProLeuGlyValGluAlaAlaAs 720

    287 .............GGGACGGTCCGGGTGGAAGGCGTGCGGGGCGTCGTGA 323
                    |||   |||         |||||||||
    720 pValGlyAlaProGlyProValProGlyProGlyValArgValArgArgA 737

    324 TGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCC 373
             |||||||||   |||:::         :::|||:::|||:::::::
    737 laGluAlaValGlyGluGlyGly.........GluGlnArgArgGluAla 750

    374 GCGGGGCAG......CCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGA 417
        |||:::          |||      |||   |||   |||   |||
    751 AlaAlaAlaArgValProGlyArgAlaArgGlyAlaLeuGlyGlyLeuGl 767

    418 ACCTCCACAACGTGTAGGGCACCGGCGGGTGGTAGCGCAGCGCCTCCTCC 467
                    ||||||   :::|||||||||   ::::::      |
    767 yAlaGluLeuLeuValGly...GlnArgValValGluHisHisHisAlaH 783

    468 ACGAACCGGGAGACGGCGTCCTC.................GTCGGCATC 499
        ||      |||   |||   |||                 ::::::
    783 isValLeuGlyValGlyTyrLeuProHisProGlyGlyAlaAlaAlaGlu 799

    500 CGCCGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGG 549
        |||      |||||||||   |||:::|||:::
    800 ArgGlyAlaAlaAlaArgGlyAspValArgGln................. 810

    550 AGCCGGTGGTGTCGTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACG 599
            ||||||   |||   ||||||   :::      :::      ::::
    811 ....GlyGlyArgValGluGlyGluArgArgAlaProGluPheGlyGluA 826

    600 AGCTGGTC..............................GTCGGAGAC 616
        ::|||:::                             ||||||
    826 spLeuLeuValHisGluGlyAlaGlyHisLeuGlyArgAlaValGlyGly 842

    617 CGAGCCGAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCAT..... 661
        ::::::::::||||||          |||      |||||||:::
    843 GluGlyArgLeuGlyGlyProArgArgValGlyLeuAlaGlyArgAspAl 859

    662 ..............GGTCGGGGTGTCGTTCCGCCGGGCTGAGTGCACG 695
                    ||||||||||||:::      |||      ::::
    859 aAlaGluAlaAlaValGlyArgGlyValLeuGlyHisGlyProGluArgA 876

    696 GCTTCGAGGAGCAGGCTCTC.........................CAG 718
        ::   |||      :::|||                         :::

    876 laProGluProValValLeuGlyGlyGlyGlyGlyGlyGlyGlnGlnArg 892

    719 GGCCGAGGTGTCCGGCAC...........................GCCCC 741
        |||   |||||||||                              |||||
    893 GlySerGlyValArgSerGlyProGluSerGluGlyAlaAlaLeuAlaPr 909

    742 CGGC.....................AGGGTCCGTGCCGTCACCCCCGCCG 770
        ||||                        |||   :::::::::||||||
    909 oGlyProProValLeuPheValValAlaValAlaValAlaValProAlaG 926

    771 CTCTGCGGGCCGCCGAGGCCGAGTGCCTTGAGAACGCTGACGGCCTCGCG 820
             ||||||:::|||          |||      :::   |||   |||
    926 luGlyArgAlaGlyGluProLeuValLeuLeuAlaValProGlyAlaAla 942

    821 GGCCATCGCCGGATCGGTGACCGGCACACCGAGCAGCTCGCAGATGACCA 870
        |||     |||
    943 GlyProGlyArg..................................... 946

    871 ACAGCGGGAAGTGGTACGCGAAGCCGCC.........GATCAGCTCGGCC 911
                                |||||                |||||||
    947 .....................AlaAlaLeuLeuLeuAlaProLeuGlyA 956

    912 GGTTTGCCCGACCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCC 961
        ||:::::|||   :::||||||            |||          |||
    956 rgTrpValArgAlaGlyGlyGly............GlyAlaGlyValAla 968

    962 GGCGATCGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGA 1011
        |||      |||:::|||         ||||||      ||| |||||||:
    969 GlyGlyAlaGlyGluAla......GlyLeuGlyAlaGlyAla.GlyLeuG 983

   1012 ACGCAGGTGCGTGGATGCGGCGCA...GGCGCCGGTGGGCCT........ 1050
        ::|||||||| |||   ||||||   |||||||||||||||
    983 lyAlaGlyAla.GlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGl 999

   1051 ...................CGCCGTCCACGGCGATGAGCGTGAACGGACG 1081
                            ||||||   ||||||            |||
    999 yGluAlaGlyGlyGlyAlaArgArgArgArgArgArgArgTrpAspAspG 1016

   1082 CAGCT............CCGGAACGGGGATGTCGAGACCGT......CGT 1113
          |||          ||||||||||||        |||      |||
   1016 luAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGlyArgGlyLeuArg 1032

   1114 CCACCCCCCGCCAGGCGGCGGGGGCGAGGTCGGGGTCCTTCACGAACCGG 1163
            |||   |||   ||||||               |||   |||||||
   1033 .GlyProGlyProArgGlyGly..............LeuGlyGluProG 1044

   1164 GGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGATGACCCAGGC 1213
        ||   ||:::          ||||||          |||      ||||||
   1044 lyArgGlyHisValGlyArgGlyGluGluGlyArgGlyValGlyProGly 1060

   1214 GGGTCCGCCCGCGGGGGC...GTTCACCTCGACGACCGGGCCCGCCTCCC 1260
        |||   |||   :::|||   ||||||
   1061 GlyLeuAlaGlyAlaGlyProValHis..................... 1069

   1261 GGAAGGCGTCGTGCACCTCGGGCTTGCGCTGCATGGTCATCATGGGACAC 1310
                             ::::::|||   :::   :::|||||||:::
   1070 ..................AlaValAlaHisGlnArgArgHisGlyAlaG 1080

   1311 GCGAACGGGTCGACGGCCACCCGGGGCGCCTCGCCGCTCACGAGGCACCG 1360
          ::::::   |||    :::||||||      ||||||   |||
   1080 lyAspGluGlyAsp.ArgValArgGlyLeuProProLeuGlyArgAlaGl 1096

   1361 CCCGCCGCCGCGGGGTACCCCTCCCGCAGTTCGACCACCGAGAAGCCGG. 1409
         |||         |||         |||   |||      |||:::|||
   1096 yPro.........GlyAspArgValAlaGluArgGluGlnArgGlyArgH 1110

   1410 ....................CCCCGTGCGGGTCGA...GCAGGTCCGCCC 1436
                            |||:::::|||||   ||||||
   1110 isLeuLeuGluAlaGlyGlyProGluGlyGlyArgGlyAlaGlyGlyArg 1126

   1437 GCCGCCCCCTGGGCGTGTCGGCGGGCTCGTTCTCGACGGAGCCGCCGAGT 1486
        :::    |||   |||:::|||      :::|||:::   ||||||
   1127 Gly.GlnProGluArgAlaGlyGlnGlnAlaLeuGluAspAlaAlaAlaG 1143
```


1487 TCAACGGCGCGCCGGACCGTCGCGTC.....GCAGTCGTGCAC 1524
1143 lyGlnAspAlaGlyValArgGlnLeuAlaGlyHisAlaAlaGlyLeuArg 1159
1525 GGCGAACAGCACGGCCCAGTGCGGCCGTACCGCGCCGGTGACGCCCAGCT 1574
1160 GlyGlyGluGlyGlyAlaAspAlaGlyAlaGluGlyLeuAspGlyArgLe 1176
1575 CCTGGGTGCCGGCGACCGGTGTGTCACCGATGTGCCAGACCGGGTCGGTG 1624
1176 uProGlyAlaGlyValArgGlyAlaAla.....ArgValGlyH 1189
1625 ACGCCCTTCAGTCCGGTGTCGGCCGGAGCCAGGCC.....GAGGGT 1665
189 isVal.....GlyValGlyProAlaGluAlaLeuGlnAspGluGly 1202
666 .....CGCCGGGT 1673
1203 LeuLeuGlyAlaIleValAlaAlaAlaHisGlyHisGlyAlaHisArgVa 1219
1674 AGAAGTCCCGGGCGGCCCCGATGCCGTC.....GGTCACCAGCTC...G 1714
1219 lArg.....GlnGlyProGluArgValLeuGlyGlyHisGlyValProA 1234
1715 ACCCAGCCGACCGAGCCGGGCACGCCCGTCACCTCCGCGCCCTCCATGAC 1764
1234 spValArgGlnArgArgGlyHisAla..... 1242
1765 TCCCTTGCGCCAGACCGCGAACGCGGCCCCGGCGGGGTCGGCGAAGACCG 1814
1243 .....AlaGlyAspGlu.....GlyAlaValAlaValGlyAr 1253
1815 C...CATCCGGCCGAGGCCGAGGACGTCCATCGGAGTCATGATGACCTCG 1861
1253 gValAspProAlaLeuAlaGluLeuValGluAlaLeuValGlyGlyLeuH 1270
1862 CCGCCCGCCGTCTCGACCCGCTTGGTCAGTGCGTCGGCGTCGTCGGTGGC 1911
1270 isProArgValLeuGlnPro...GlyHis.....GlyLeuAlaGlyGly 1283
1911 1911
1284 ProAlaValAspGluAlaHisGluGlyLeuValLeuLeuProArgValPr 1300
1912 .....GAAGTACACGGTCCAGATGGCCGGCATGCC.... 1941
1300 oHisLeuArgAspGluAspGlyHisGlyPro...GlyArgGlyAlaValA 1316
1942 .....GTGCTGGTCGTTCCCGGGCCCGTACGGCCG 1971
1316 laGlyArgGlyLeuAlaAspValValLeuValProGluProLeuAlaGly 1332
1972 GTGGTAGGGGGT.....GTCGATCTGGTGGCGGGC..... 2001
1333 ValProGlyAlaAlaValAlaAspAlaAlaValGlyArgArgValGlyAl 1349
2002 .....GACCGCGGCGACCAGCTTCCCGTCGGAGCT.... 2031
1349 aGlyProGlyLeuProGluArgGlyGluGlnArgProValGlyArgArgG 1366
2032 .....GAACGTCGT...GTATCCCCCGGCGCCCGGGTCGCT 2064
1366 lyProValGlyHisGluArgGluValValValGlyGlyAlaAlaLeuPro 1382
2065 GACCACGGTGGCGGTCCAGCCGAACAGGCCGGTGTAGAAGTCGGCCGAGG 2114
1383 AlaArgGlyProGlyGlyLeuArgGlyArgGly...ArgGlyGlyArgGl 1398
2115 CGGCGACAT.....CGGGCGAACCGAGGTCGAACCATGCGGG 2151
1398 yGlyGlyGlyGlyGlyGlyGlyArgGlyProArgGlyArg.....GlyG 1413
2152 GGCGCCGGGCGCGAACCTGGTCACGAATCGTTCCTTTCGATGGATCGGCA 2201
1413 lyArgArgArgArgArgTrpArgProGlyAlaGlyGluTrpGlyAlaGly 1429
2202 CACGAGCGTCTG.....CGCTCGCGGATGAG 2227
1430 ProAspSerPheValPhePheSerLeuGlyGlyGlyArgGlyArgGlyGl 1446
2228 ACGGACATCTCGCGGATGAGACGGACATGCGGGCGGGGC.GGGCCGCCGC 2276
1446 yArgGlyGlyArgGlyGlyArgGlyArgGlyGlyGlyArgAlaProArgG 1463
2277 CGTCAGTGCGCGGTGTCGCCGACGGCGGCCGCGCCGGCCTCCCAGAGCTT 2326
1463 lyGlyGlyGlyGlyProGlyGlyGlyGlyArgAlaGlyArgGlyGluVal 1479
2327 CGCCGCGAGGCCGGCGTCGGCGGTCGGGCCGCTCACCGGGGACAGCCGCC 2376
1480 ArgValAlaAlaAlaAlaAlaGlyAlaAlaGlu..... 1490
2377 GGTCGCTGTAGTAGCCGCCCGTGGTCAACTCCTCGGCCGG..... 2416
1491 ....AlaAlaAlaAlaAlaGluGlyAlaLeuSerGlyAlaAlaProAlaG 1506
2417 .....CGCGGACGCCAG.....CCACACGAG..... 2437
1506 lyLeuSerLeuArgGlyArgProAlaValProGlyGluAlaGluSerVal 1522
2437 2437
1523 LeuValLeuLeuGlyAlaAlaGlyAspGlyLeuAspGlyAspGlyGlyGl 1539
2438 .....GGTGTCGGCGCCCTTCGCCGCGGAGCG..... 2464
1539 yGlyValGlyValGlyValGlyValGlyGlyAspGlyAlaProGlyAlaL 1556
2465 ..CAGGAAGGGGTTGAACCGGAAGTAGGACGAGGC..... 2497
1556 ysArgProArgIleGluProProArgGlyGlyGlyLeuValGluGlnGly 1572
2498 .....GACCGTGCCCCGTCCGATGCGGGTGC 2523
1573 LeuAlaValLeuValMetValThrThrAlaValProSerGlyGlyGly.. 1588
2524 GGACCTCACCGGGGTGATAGCTGACCGCCAGCACGTCCGGCCAGCGCCTG 2573
1589 .....GlyAlaAlaAlaAlaGlyArgArgAspArgProGlyGlyG 1602
2574 GCGGC.....CTCCGCCGCGGTCATGA. 2595
1602 lyGlyGlyTrpGlySerGlyProProProCysArgArgCysGlyHisArg 1618
2596 TGTTGGCCTGTTTGGACGTGCCGTACGCCT..... 2625
1619 CysTrpLeuCysTrpTrpArgArgGlyProArgProArgArgArgProGl 1635
2626 .....GGCCGGCGCTGTA 2638
1635 yLeuThrAspArgValProProArgGlyGlyProSerProArgGlyCysA 1652
2639 GCGGTGACGGTCGCCGTTGAGGT..... 2661
1652 rgGlyAlaGlyArgAlaGlyGlyGlyGlyArgGlyGlyCysGlyGlyGly 1668
2662 ...CGTCCGGGTCGATCC...GGCCCTGGG.....TGTAC 2690
1669 ArgAlaProGlyAlaAlaGlyGlyProGlyLeuCysArgCysGluCysCy 1685
2691 GCGTCGGACGAGGTGAGGATCAGCCGCCCGCCCGCGAGCCGCTCCCGCAG 2740
1685 sArgGlyArgArgProGlyProGlyAlaGlyProGlyProGlyProGluA 1702
2741 CAGCCGTG.....CCAGCAGGAAGCCTG 2763

```
          |||                          |||:::|||:::
    1702 spGluValThrValLeuGlyAlaLeuMetGluSerProThrGlyGlyPro 1718

    2764 CGAGGTGATTGACCTGGA 2781
           |||      ||||||
    1719 GlyGlyArgGlyProGly 1724

seq_name: pir1:A35175

seq_documentation_block:
mucin 1 precursor, repetitive splice form A [validated] - human
N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin
ncreatic mucin; polymorphic epithelial mucin (PEM)
N;Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C;Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A;Reference number: A35175; MUID:90202794
A;Accession: A35175
A;Molecule type: mRNA
A;Residues: 1-952,1033-1344 <LIG1>
A;Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:
A;Experimental source: splice form A
A;Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl-en
A;Accession: B35175
A;Molecule type: mRNA
A;Residues: 1-19,29-952,1033-1344 <LIG2>
A;Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:
A;Experimental source: splice form B
A;Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-en
R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel
J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli
A;Reference number: A35886; MUID:90368715
A;Accession: A35886
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-992,1033-1344 <GEN>
A;Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
A;Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A;Reference number: A35887; MUID:90368716
A;Accession: A35887
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A;Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
A;Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated seque
R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A;Reference number: S10571; MUID:90276413
A;Accession: S10572
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A;Cross-references: EMBL:X52229; NID:g37053
R;Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A;Reference number: S40293
A;Accession: S40293
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2
A;Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
R;Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A;Reference number: A36735; MUID:90088473
A;Accession: A36735
A;Molecule type: mRNA
A;Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A;Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu
J. Biochem. 112, 609-615, 1992
A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut ag
A;Reference number: JX0235; MUID:93123189
A;Accession: PX0066
A;Molecule type: mRNA
A;Residues: 998-1011,'ES',1014-1017;1018-1032,'T',1034-1037;1038-1057 <MAS>
A;Experimental source: gastric carcinoma cell
R;Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokin
A;Reference number: S51026; MUID:95080414
A;Contents: annotation
A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region
C;Comment: This protein is length polymorphic. Individuals may have between 21 and 12
 partial repeats. The repeat shown is defined by SmaI nuclease sites.
C;Comment: Serine and threonine residues in the tandem repeat domain are extensively
C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48
C;Genetics:
A;Gene: GDB:MUC1; PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map position: 1q21-1q23
A;Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C;Superfamily: polymorphic epithelial mucin
C;Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorp
F;1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F;1-62/Region: mucin 1 amino-terminal non-repetitive
F;1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F;1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F;1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F;1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice f
F;138-1017/Region: 20-residue repeats (GSTAPPAHGVTSAPDTRPAP)
F;1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F;1245-1272/Domain: transmembrane #status predicted <TRM>
F;1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predic
F;1213/Binding site: phosphate (Tyr) (covalent) #status predicted


alignment_scores:
           Quality:  464.50              Length:   1065
             Ratio:  0.950                 Gaps:     62
 Percent Similarity:  45.915   Percent Identity: 28.263

alignment_block:
 US-09-673-254-1/rev x A35175   ..

 Align seg 1/1  to: A35175  from: 1  to: 1344

    2865 ACCCGCGCATCGATGTCATGGCCAGCAACGCCGGCGGCATGTTCTGGTCG 2816
         ||||||  :.........      :::   ||||||
     113 ThrArgProAlaLeuGlySerThrThrProProAlaHisAspValThrSe 129

    2815 CGCACCA...:..CGACCCAGGACGGGTTCG...AGGCCACCATCCAGGT 2775
          ||||||      :::|||      ||||||      ||||||:::::::
     129 rAlaProAspAsnLysProAlaProGlySerThrAlaProProAlaHisG 146

    2774 CAATCACCTCGCAGGCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCG 2725
         :::::||||||      :::         :::      :::::::::
     146 lyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 162

    2724 CGGGCGGGCGGCTGATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGG 2675
            |||   |||   :::|||        ||||||   |||      ||
     163 ProAlaHisGlyValThrSerAlaProAspThrArg...ProAlaProGl 178

    2674 ATCGACCCGGACGACCTCAACGGCGACCGTCACCGCT......ACAGCGC 2631
         |||||||         :::   :::::::::::::|||      |||   |
     178 ySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgP 195

    2630 CGGCCAGGCGTACGGCACGTCCA......AACAGGCCAACATCATGACCG 2587
         |||||      ::::::   |||      :::      ||||||   |||
```


195 roAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaPro 211
2586 CGGCGGAGGCCGCCAGGCGCTGGCCGGACGTGCTGGCGGTCAGCTATCAC 2537
212 AspThrArgPro....Al 216
2536 CCCGGTGAGGTCCGCACCCGCATCGGACGGGGCACGGTCG.....C 2496
216 aProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspT 233
2495 CTCGTCCTACTTCCGGTTCAACCCCTTCCTGCGCTCCGCGGCGAAGGGCG 2446
233 hrArgProAlaProGlySerThr.....AlaProProAlaHisGly 246
2445 CCGACACCCTCGTGTGGCTGGCGTCCGCGCCGGCCGAGGAGTTGACCACG 2396
247 ValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPr 263
2395 GGCG.....GCTACTACAGCGACCGGCGGCTGTCCCCGGTGAGCGGCCC 2352
263 oAlaHisGlyValThrSerAla..... 270
2351 GACCGCCGACGCCGGCCTCGCGGCGAAGCTCTGGGAGGCCGGCGCGGCCG 2302
271 ..ProAspThrArgProAlaProGlySerThrAlaProProAlaHisGly 286
2301 CCGTCGGCGACACCGCGCACT.....GACGGCGGCGGCCCGCCCCG 2261
287 ValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPr 303
2260 CCCGCATGTCCGTCTCATCCGCGAGATGTCCGTCTCATCCGCGAGCGCAG 2211
303 oAlaHis.GlyValThrSerAla.....ProAspThrArgProAla... 316
2210 ACGCTCGTGTGCCGATCCATCGAAAGGAACGATTCGTGACCAGGTTCGCG 2161
317 .....ProGlySerTh 320
2160 CCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCCCGATGTCGCCGCCTC 2111
320 rAlaProProAlaHisGlyValThrSerAla...ProAspThrArgProA 336
2110 GGCCGACTTCTACACCGGCCTGTTCGGCTGGACCGCCACCGTGGTCAGCG 2061
336 laProGlySerThr.....AlaProProAlaHisGlyVal 347
2060 ACCCGGGCGCCGGGGGATACACGA.....CGTTCAGCTCCGAC 2023
348 ThrSerAlaPro...AspThrArgProAlaProGlySerThrAlaProPr 363
2022 GGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCCCTACCA 1973
363 oAlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerT 380
1972 CCGGCCGTACGGGCCCGGGAACGACCAGCACGGCA...TGCCGGCCATCT 1926
380 hrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAla 396
1925 GGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGGGTC 1876
397 ProGlySerThrAlaProProAlaHisGlyVal.....ThrSerAlaPr 411
1875 GAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGGCCT 1826
411 oAspThrArgProAlaProGlySer.....ThrAlaProProA 424
1825 CGGCCGGATGGCGGTCTTCGCCG...ACCCCGCCGGGGCCGCGTTCGCGG 1779
424 laHisGlyValThrSerAlaProAspThrArgProAlaProGlySer... 439
1778 TCTGGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGC 1729
440 .....ThrAlaProProAl 444
1728 TCGGTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCGGGA 1679
444 aHisGlyValThrSerAla.....ProAspThrArgProAlaProGly. 458
1678 CTTCTACCCGGCGACCCTCGGCCTGGCTC.....CGG 1647
459 ..SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 474
1646 CCGACACCGGACTGAAGGGCGTCACCGACCCGGTCTGGCACATCGGTGAC 1597
475 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaPr 491
1596 ACACCGGTCGCCGGCACCCAGGAGCTGGGCGTCACCGGCGCGGTACGGCC 1547
491 oAspThrArgProAlaProGlySerThrAlaProProAlaHisGlyValT 508
1546 GCACTGGGCCGTGCTGTTCGCCGTGCACGACTGCGACGCGACGG..... 1503
508 hrSerAlaProAspThrArgProAlaProGlySerThrAlaProProAla 524
1502 .....TCCGGCGCGCCGTTGAACTCGGCGGCTCCGTCGAGAACGAG 1462
525 HisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAl 541
1461 CCCGCCG.....ACACGCCCAGGGGGCG 1439
541 aProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProG 558
1438 GC.....GGGCGGACCTGCTCGACCCGCACGGGG 1410
558 lySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 574
1409 CCGGCTTCTCGGTGGTCGAACTGCGGGAGGGGTACCCCGCGGCGGCGGGC 1360
575 ProAlaPro.....GlySerThrAlaProProAlaHisGl 586
1359 GGTGCCTCGT.....GAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCC. 1317
586 yValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProP 603
1316 .....GTTCGCGTGTCCCATGATGAC 1296
603 roAlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySer 619
1295 CATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGG 1246
620 ThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAl 636
1245 TCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATCACCGAT 1196
636 aProGlySer..ThrAlaProProAlaHisGlyValThrSerAlaProAs 652
1195 GACG...CCCTCGCCCGCGAGGTGCTGGCCG..... 1168
652 pThrArgProAlaProGlySerThrAlaProProAlaHisGlyValThrS 669
1167 ..ATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCGCCGCCTGGCGGG 1120
669 erAlaProAspThrArgProAlaProGlySerThrAlaProProAlaHis 685
1119 GGGTGGACGACGGTC.....TCGACATCCCC 1094
686 GlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPr 702
1093 G.....TTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGA 1056
702 oProAlaHisGlyValThrSerAlaProAspThrArgProAlaPro.Gly 718
1055 CGGCGAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACC 1006
719 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgPr 735



OM of: US-09-673-254-1 to: SwissProt_40:* out_format : pfs

Date: Jun 11, 2002 8:21 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09673254/runat_11062002_114213_1053/app_query.fasta_1.2973
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
[illegible]LOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
[illegible]NS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
[illegible]_MAX=100 -THR_MIN=0 -ALIGN=7 -MODE=LOCAL -OUTFMT=pfs
[illegible]M=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09673254_@CGN1_1_80 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-673-254-1
Query length: 2870
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 57.220000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
SwissProt_40:VNUA_PRVKA	+	468.50	285.30	1.9e-09	1733	P33485 pseudorabies virus (
SwissProt_40:CA11_CHICK	-	459.50	281.14	4.0e-09	1453	P02457 gallus gallus (chick
SwissProt_40:CA11_CANFA	-	459.50	281.11	4.0e-09	1460	Q9xsj7 canis familiaris (dc
SwissProt_40:CA11_HUMAN	-	457.00	279.62	4.8e-09	1464	P02452 homo sapiens (human)
SwissProt_40:MUC1_HUMAN	-	455.50	279.73	5.5e-09	1255	P15941 h mucin 1 precursor
SwissProt_40:CA11_MOUSE	-	450.00	275.55	8.1e-09	1453	P11087 mus musculus (mouse)
SwissProt_40:CA13_MOUSE	-	450.00	275.50	8.1e-09	1464	P08121 mus musculus (mouse)
SwissProt_40:MUC2_HUMAN	-	449.50	267.09	6.7e-09	5179	Q02817 homo sapiens (human)
SwissProt_40:YHL1_EBV	-	442.00	275.91	1.7e-08	660	P03181 epstein-barr virus (str
SwissProt_40:CA13_BOVIN	-	442.00	272.94	1.6e-08	1049	P04258 bos taurus (bovine).
SwissProt_40:CA21_RANCA	-	439.00	269.53	1.9e-08	1355	O42350 rana catesbeiana (bu
SwissProt_40:CA25_HUMAN	-	439.00	268.89	1.8e-08	1496	P05997 homo sapiens (human)
[illegible]SwissProt_40:CA13_CHICK	-	437.50	269.10	2.1e-08	1262	P12105 gallus gallus (chick
[illegible]SwissProt_40:VNUA_PRVKA	-	437.50	267.07	2.0e-08	1733	P33485 pseudorabies virus (
[illegible]SwissProt_40:CF30_MYCTU	-	434.50	277.49	3.5e-08	260	O53774 mycobacterium tubercu
[illegible]SwissProt_40:CA13_HUMAN	-	432.50	265.20	3.0e-08	1466	P02461 homo sapiens (human)
SwissProt_40:CA11_HUMAN	+	418.50	256.98	8.7e-08	1464	P02452 homo sapiens (human)
SwissProt_40:CA17_HUMAN	+	417.00	251.61	8.6e-08	2944	Q02388 homo sapiens (human)
SwissProt_40:CA15_HUMAN	+	414.50	253.16	1.1e-07	1838	P20908 homo sapiens (human)
SwissProt_40:CA12_MOUSE	-	413.50	254.06	1.3e-07	1459	P28481 mus musculus (mouse)
SwissProt_40:CA13_HUMAN	+	413.50	254.03	1.3e-07	1466	P02461 homo sapiens (human)
SwissProt_40:CA21_ONCMY	-	412.00	253.65	1.4e-07	1356	O93484 oncorhynchus mykiss
SwissProt_40:CPXK_SACER	-	411.50	261.12	1.9e-07	405	P33271 saccharopolyspora ery
SwissProt_40:CA17_HUMAN	-	410.50	247.78	1.4e-07	2944	Q02388 homo sapiens (human)
SwissProt_40:CA12_HUMAN	-	409.50	251.89	1.7e-07	1418	P02458 homo sapiens (human)
SwissProt_40:CA11_CANFA	+	409.50	251.70	1.7e-07	1460	Q9xsj7 canis familiaris (dc
SwissProt_40:AMYH_YEAST	-	408.00	251.24	1.9e-07	1367	P08640 saccharomyces cerevi
SwissProt_40:CPXH_STRGR	-	406.50	258.07	2.7e-07	412	P26911 streptomyces griseus.
SwissProt_40:CA21_HUMAN	-	406.00	250.07	2.3e-07	1366	P08123 homo sapiens (human)
SwissProt_40:CA13_MOUSE	+	405.00	249.04	2.4e-07	1464	P08121 mus musculus (mouse)
SwissProt_40:CA13_BOVIN	+	403.00	250.00	3.0e-07	1049	P04258 bos taurus (bovine).
SwissProt_40:CA15_HUMAN	-	403.00	246.40	2.7e-07	1838	P20908 homo sapiens (human)
SwissProt_40:YHL1_EBV	+	402.50	252.69	3.4e-07	660	P03181 epstein-barr virus (str
SwissProt_40:CA21_MOUSE	+	402.50	247.98	2.9e-07	1372	Q01149 mus musculus (mouse)
SwissProt_40:CPXP_RHISN	-	401.50	255.31	3.9e-07	400	P55544 rhizobium sp. (strain
SwissProt_40:CA21_RAT	+	399.00	245.93	3.8e-07	1372	P02466 rattus norvegicus (rat
SwissProt_40:CA11_MOUSE	+	398.00	244.97	4.1e-07	1453	P11087 mus musculus (mouse)
SwissProt_40:CA21_CANFA	-	397.00	244.78	4.5e-07	1366	O46392 canis familiaris (dc
SwissProt_40:CA1B_HUMAN	-	397.00	242.98	4.3e-07	1806	P12107 homo sapiens (human)
SwissProt_40:CA11_CHICK	+	396.50	244.09	4.6e-07	1453	P02457 gallus gallus (chick
SwissProt_40:CA21_HUMAN	+	396.00	244.19	4.8e-07	1366	P08123 homo sapiens (hum
SwissProt_40:CA1B_HUMAN	+	395.00	241.81	4.9e-07	1806	P12107 homo sapiens (hum
SwissProt_40:CA11_BOVIN	-	394.50	246.92	6.0e-07	779	P02453 bos taurus (bovine
SwissProt_40:CA21_BOVIN	-	394.50	243.32	5.4e-07	1364	P02465 bos taurus (bovin
SwissProt_40:CA25_HUMAN	+	394.50	242.72	5.3e-07	1496	P05997 homo sapiens (hum

seq_name: SwissProt_40:VNUA_PRVKA

seq_documentation_block:
ID VNUA_PRVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vlcek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M34651; AAA47471.1; -.
DR PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

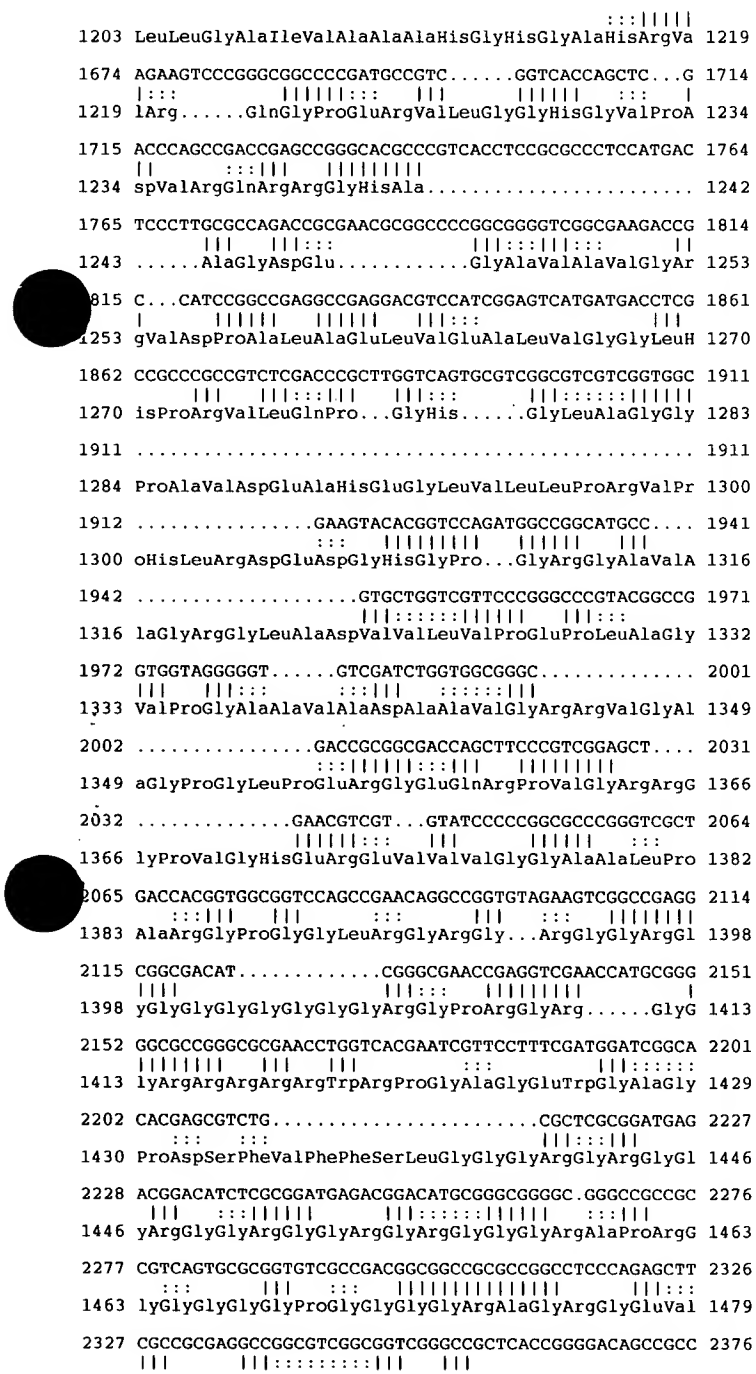
alignment_scores:
Quality: 469.00 Length: 1226
Ratio: 0.925 Gaps: 72
Percent Similarity: 41.354 Percent Identity: 29.364

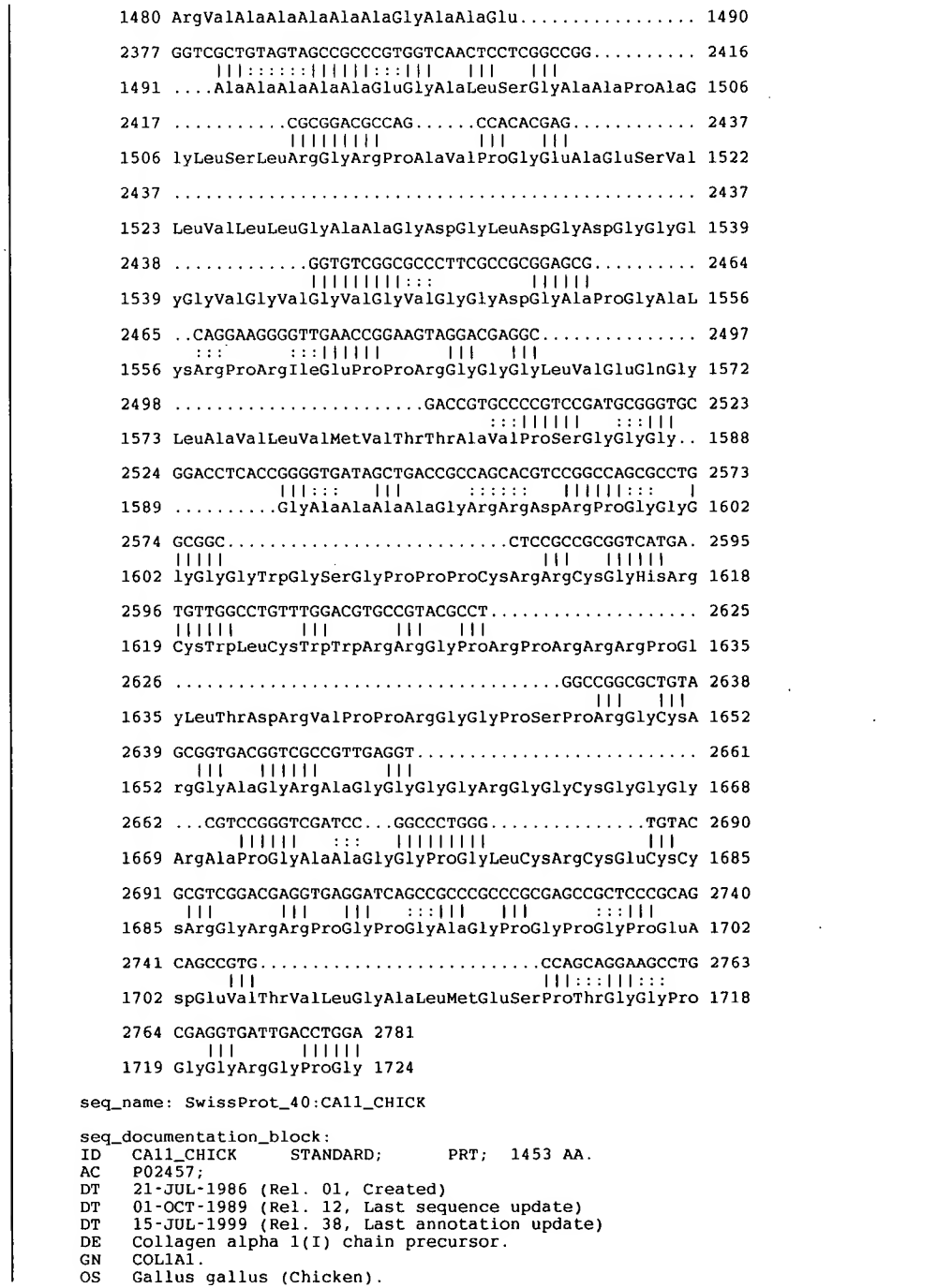
alignment_block:
US-09-673-254-1 x VNUA_PRVKA ..

Align seg 1/1 to: VNUA_PRVKA from: 1 to: 1733

```
      2 GATCCGCACCGGGTA...CACGGCACGGGACCGCCCACCGCGCGGTGCGC 48
        |||      ||||||  ||||||              |||:::  ||
    631 AspGlyAlaArgValGlnHisGly.............ArgAlaGluLeuAl 643

     49 GGTGGGCGGTCCCGTGCCGGTCGCGGCCGGCGGATCAGCGCAGCCAGACG 98
        |:::  ||||||:::     |||  ||||||  :::|||  |||
    643 aAlaProGlyProAlaHisGlyAlaLeuGlyGlyArgValGlnAlaAsp. 659

     99 GGCAGTTCGGTGAGCCGCGCCGTCTGGGCCCCCTTCCGGCACCACCGCAA 148
                                          |||     :::
    660 ....................................AlaAspValAsp 663

    149 CTCGTCGTACGGCACGGCCAGTCGGGCCTCGGG..................  181
        :::||||||  ||||||::::::     |||
    664 ValValValProHisGlyArgAlaValArgGlyProValLeuAspGlyVa 680
```

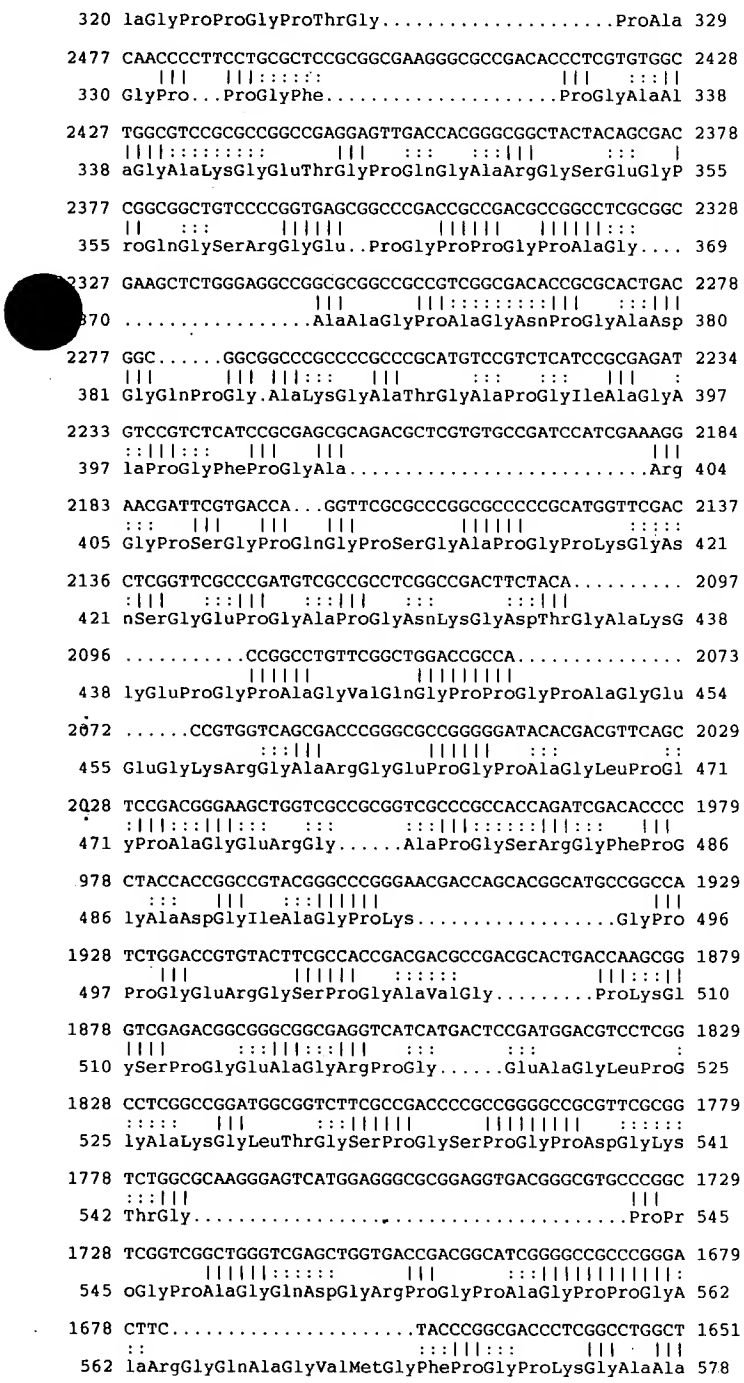
```
182 .........GAACCTGCTGCGCAGTACGCCGATCATCGTGCGCGACTCC 221
680 lGlnHisAspGluProAlaProArgArgAlaGluProArgAlaGlu.... 695

222 AGCTGGGCGAGCTGCTCCCCGATGCAGTAGTGCGGCCCGTCGCCGAAGGT 271
696 .........................ValLeuHisGlyAlaGlyGluAla 703

272 GAGCCGCCGCCACGA................................. 286
704 GluValProArgArgGluGlnGlnHisProLeuGlyValGluAlaAlaAs 720

287 .............GGGACGGTCCGGGTGGAAGGCGTGCGGGGCGTCGTGA 323
720 pValGlyAlaProGlyProValProGlyProGlyValArgValArgArgA 737

324 TGGCGGCCGTCGGTGTTGGTGCCCTCGATGTCCACCAGCACCGGCGCTCC 373
737 laGluAlaValGlyGluGlyGly.........GluGlnArgArgGluAla 750

374 GCGGGGCAG......CCGGACGCCGCCGATGGTCACCTCCGTGGCAGCGA 417
751 AlaAlaAlaArgValProGlyArgAlaArgGlyAlaLeuGlyGlyLeuGl 767

418 ACCTCCACAACGTGTAGGGCACCGGCGGGTGGTAGCGCAGCGCCTCCTCC 467
767 yAlaGluLeuLeuValGly...GlnArgValValGluHisHisHisAlaH 783

468 ACGAACCGGGAGACGGCGTCCTC..................GTCGGCATC 499
783 isValLeuGlyValGlyTyrLeuProHisProGlyGlyAlaAlaAlaGlu 799

500 CGCCGCGAGGCGGCCCGCCAGGACCTCCGCGAGCAGGAAGCCCAGGAAGG 549
800 ArgGlyAlaAlaAlaArgGlyAspValArgGln................... 810

550 AGCCGGTGGTGTCGTGGCCGGCGAAGATGAGCCCGGTGATCATGTAGACG 599
811 ....GlyGlyArgValGluGlyGluArgArgAlaProGluPheGlyGluA 826

600 AGCTGGTC................................GTCGGAGAC 616
826 spLeuLeuValHisGluGlyAlaGlyHisLeuGlyArgAlaValGlyGly 842

617 CGAGCCGAACTCGGCCTGCGCGCGCTCGTACAGCACGCGGGTCAT..... 661
843 GluGlyArgLeuGlyGlyProArgArgValGlyLeuAlaGlyArgAspAl 859

662 ................GGTCGGGGTGTCGTTCCGCCGGGCTGAGTGCACG 695
859 aAlaGluAlaAlaValGlyArgGlyValLeuGlyHisGlyProGluArgA 876

696 GCTTCGAGGAGCAGGCTCTC.........................CAG 718
876 laProGluProValValLeuGlyGlyGlyGlyGlyGlyGlyGlnGlnArg 892

719 GGCCGAGGTGTCCGGCAC.........................GCCCC 741
893 GlySerGlyValArgSerGlyProGluSerGluGlyAlaAlaLeuAlaPr 909

742 CGGC....................AGGGTCCGTGCCGTCACCCCCGCCG 770
909 oGlyProProValLeuPheValValAlaValAlaValAlaValProAlaG 926

771 CTCTGCGGGCCGCCGAGGCCGAGTGCCTTGAGAACGCTGACGGCCTCGCG 820
926 luGlyArgAlaGlyGluProLeuValLeuLeuAlaValProGlyAlaAla 942

821 GGCCATCGCCGGATCGGTGACCGGCACACCGAGCAGCTCGCAGATGACCA 870
943 GlyProGlyArg..................................... 946

871 ACAGCGGGAAGTGGTACGCGAAGCCGCC.........GATCAGCTCGGCC 911
947 .....................AlaAlaLeuLeuLeuAlaProLeuGlyA 956

912 GGTTTGCCCGACCGGCCGGAGGCGTCGGCGAGTTCGGTGAGCAGCCGGCC 961
956 rgTrpValArgAlaGlyGlyGly.............GlyAlaGlyValAla 968

962 GGCGATCGCGGCGATGCGATCCGTCCGCTCGGCCAGCCGGCGCGGGTTGA 1011
969 GlyGlyAlaGlyGluAla......GlyLeuGlyAlaGlyAla.GlyLeuG 983

1012 ACGCAGGTGCGTGGATGCGGCGCA...GGCGCCGGTGGGCCT........ 1050
983 lyAlaGlyAla.GlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGl 999

1051 ...................CGCCGTCCACGGCGATGAGCGTGAACGGACG 1081
999 yGluAlaGlyGlyGlyAlaArgArgArgArgArgArgTrpAspAspG 1016

1082 CAGCT.............CCGGAACGGGGATGTCGAGACCGT......CGT 1113
1016 luAlaGlyLeuGluGlyProGluArgGlyGlnAlaGlyArgGlyLeuArg 1032

1114 CCACCCCCCGCCAGGCGGCGGGGGCGAGGTCGGGGTCCTTCACGAACCGG 1163
1033 .GlyProGlyProArgGlyGly..............LeuGlyGluProG 1044

1164 GGATCGGCCAGCACCTCGCGGGCGAGGGCGTCATCGGTGATGACCCAGGC 1213
1044 lyArgGlyHisValGlyArgGlyGluGluGlyArgGlyValGlyProGly 1060

1214 GGGTCCGCCCGCGGGGGC...GTTCACCTCGACGACCGGGCCCGCCTCCC 1260
1061 GlyLeuAlaGlyAlaGlyProValHis....................... 1069

1261 GGAAGGCGTCGTGCACCTCGGGCTTGCGCTGCATGGTCATCATGGGACAC 1310
1070 ...................AlaValAlaHisGlnArgArgHisGlyAlaG 1080

1311 GCGAACGGGTCGACGGCCACCCGGGGCGCCTCGCCGCTCACGAGGCACCG 1360
1080 lyAspGluGlyAsp.ArgValArgGlyLeuProProLeuGlyArgAlaGl 1096

1361 CCCGCCGCCGCGGGGTACCCCTCCCGCAGTTCGACCACCGAGAAGCCGG. 1409
1096 yPro.........GlyAspArgValAlaGluArgGluGlnArgGlyArgH 1110

1410 ...................CCCCGTGCGGGTCGA...GCAGGTCCGCCC 1436
1110 isLeuLeuGluAlaGlyGlyProGluGlyGlyArgGlyAlaGlyGlyArg 1126

1437 GCCGCCCCCTGGGCGTGTCGGCGGGCTCGTTCTCGACGGAGCCGCCGAGT 1486
1127 Gly.GlnProGluArgAlaGlyGlnGlnAlaLeuGluAspAlaAlaAlaG 1143

1487 TCAACGGCGCGCCGGACCGTCGCGTC............GCAGTCGTGCAC 1524
1143 lyGlnAspAlaGlyValArgGlnLeuAlaGlyHisAlaAlaGlyLeuArg 1159

1525 GGCGAACAGCACGGCCCAGTGCGGCCGTACCGCGCCGGTGACGCCCAGCT 1574
1160 GlyGlyGluGlyGlyAlaAspAlaGlyAlaGluGlyLeuAspGlyArgLe 1176

1575 CCTGGGTGCCGGCGACCGGTGTGTCACCGATGTGCCAGACCGGGTCGGTG 1624
1176 uProGlyAlaGlyValArgGlyAlaAla............ArgValGlyH 1189

1625 ACGCCCTTCAGTCCGGTGTCGGCCGGAGCCAGGCC.........GAGGGT 1665
1189 isVal.........GlyValGlyProAlaGluAlaLeuGlnAspGluGly 1202

1666 .....................................CGCCGGGT 1673
```






```
seq_name: SwissProt_40:CA11_CHICK

seq_documentation_block:
ID   CA11_CHICK     STANDARD;      PRT;  1453 AA.
AC   P02457;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1.
OS   Gallus gallus (Chicken).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=88056316; PubMed=3678834;
RA   Finer M.H., Boedtker H., Doty P.;
RT   "Construction and characterization of cDNA clones encoding the 5' end
RT   of the chicken pro alpha 1(I) collagen mRNA.";
RL   Gene 56:71-78(1987).
RN   [2]
RP   SEQUENCE OF 1-144 FROM N.A.
RX   MEDLINE=88007542; PubMed=2820966;
RA   Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT   "Unusual DNA sequences located within the promoter region and the
RT   first intron of the chicken pro-alpha 1(I) collagen gene.";
RL   J. Biol. Chem. 262:13323-13332(1987).
RN   [3]
RP   SEQUENCE OF 152-1187.
RX   MEDLINE=82231995; PubMed=7093229;
RA   Highberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA   Kang A.H., Gross J.;
RT   "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
RT   complete primary structure of the helical portion of the chick skin
RT   collagen alpha 1(I) chain.";
RL   Biochemistry 21:2048-2055(1982).
RN   [4]
RP   SEQUENCE OF 1200-1205.
RX   MEDLINE=72243016; PubMed=5047697;
RA   Eyre D.R., Glimcher M.J.;
RT   "Evidence for a previously undetected sequence at the carboxyterminus
RT   of the alpha 1 chain of chicken bone collagen.";
RL   Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN   [5]
RP   SEQUENCE OF 981-1453 FROM N.A.
RX   MEDLINE=81160715; PubMed=6927845;
RA   Fuller F., Boedtker H.;
RT   "Sequence determination and analysis of the 3' region of chicken pro-
RT   alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT   including the carboxy-terminal propeptide sequences.";
RL   Biochemistry 20:996-1006(1981).
RN   [6]
RP   SEQUENCE OF 1311-1453 FROM N.A.
RX   MEDLINE=80134546; PubMed=6987088;
RA   Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA   Pastan I., Decrombrugghe B., Fietzek P.P., Olsen B.R.;
RT   "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT   carboxyl end of pro alpha 1(I)-chains.";
RL   FEBS Lett. 111:61-65(1980).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17839; AAA48704.1; -.
DR   EMBL; M17838; AAA48704.1; JOINED.
DR   EMBL; V00401; CAA23695.1; -.
DR   EMBL; M10571; AAA48671.1; ALT_SEQ.
DR   EMBL; M17607; AAA48672.1; -.
DR   PIR; A02857; CGCH1S.
DR   PIR; A27179; A27179.
DR   PIR; A29367; A29367.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     22
FT   PROPEP      23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       152   1205       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1206   1453       C-TERMINAL PROPEPTIDE.
FT   DOMAIN       31     89       VWFC.
FT   MOD_RES     152    152       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     254    254       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     851    851       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1081   1081       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1097   1097       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1153   1153       HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT                                ONLY HYDROXYLATED PRO IN POSITION X (IN
FT                                THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
FT   CONFLICT   1187   1187       F -> L (IN REF. 5).
FT   CONFLICT   1441   1441       Q -> H (IN REF. 6).
SQ   SEQUENCE   1453 AA;  137789 MW;  3BC6152134271F4D CRC64;


alignment_scores:
            Quality:  459.50              Length:   1129
              Ratio:  0.963                 Gaps:     64
 Percent Similarity:  42.250   Percent Identity: 27.015

alignment_block:
 US-09-673-254-1/rev x CA11_CHICK   ..

 Align seg 1/1  to: CA11_CHICK  from: 1  to: 1453

    2844 CCAGCAACGCCGGCGGCATGTTCT.........GGTCGCGCACCACGACC 2804
         ||||||   |||   :::   :::          |||:::|||   :::||
     211 ProAlaGlyProProGlyLysAsnGlyAspAspGlyGluAlaGlyLysPr 227

    2803 C...AGGACGGGTTCGAGGCCACCATCCAGGT...CAATCACCTCGCAGG 2760
         |   |||   |||:::|||   ||| ||||||   |||:::   |||
     227 oGlyArgProGlyGlnArgGlyPro.ProGlyProGlnGlyAlaArgGly 243

    2759 CTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGCGGGCGGCTGA 2710
         ||||||   |||||||||:::      :::   :::|||:::   :::
     244 LeuPro...GlyThrAlaGlyLeuProGlyMetLysGlyHisArgGlyPh 259

    2709 TCCTCACCTC...GTCCGACGCGTACACCCAGGGCCG............G 2675
               |||         :::   :::|||||||||
     259 eSerGlyLeuAspGlyAlaLysGlyGlnProGlyProAlaGlyProLysG 276

    2674 ATCGACCCGGACGACCTCAACGGCGACCGTCACCGCTACAGCGCCGGCCA 2625
           :::||||||   |||   ::::::   :::|||   |||       |||
     276 lyGluProGlySerProGlyGluAsnGlyAlaProGlyGlnMetGlyPro 292

    2624 GGCGTACGGCACGTCCAAACAGGCCAACATCATGACCGCGGC...GGAGG 2578
                                  |||         :::||||||      ||
     293 .................ArgGlyLeuProGlyGluArgGlyArgProGl 303

    2577 CCGCCAGGCGCTGGCCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAG 2528
         |   ::::::   ||||||      |||::::::   :::|||      :
     303 yProSerGlyProAlaGlyAlaArgGlyAsnAspGlyAlaProGlyAlaA 320

    2527 GTCCGCACCCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTT 2478
         ::      |||      ||||||                     |||:::
```


320 laGlyProProGlyProThrGly.....ProAla 329
2477 CAACCCCTTCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGC 2428
330 GlyPro...ProGlyPhe.....ProGlyAlaAl 338
2427 TGGCGTCCGCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGAC 2378
338 aGlyAlaLysGlyGluThrGlyProGlnGlyAlaArgGlySerGluGlyP 355
2377 CGGCGGCTGTCCCCGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGGC 2328
355 roGlnGlySerArgGlyGlu..ProGlyProProGlyProAlaGly.... 369
2327 GAAGCTCTGGGAGGCCGGCGCGGCCGCCGTCGGCGACACCGCGCACTGAC 2278
370 .....AlaAlaGlyProAlaGlyAsnProGlyAlaAsp 380
2277 GGC.....GGCGGCCCGCCCCGCCCGCATGTCCGTCTCATCCGCGAGAT 2234
381 GlyGlnProGly.AlaLysGlyAlaThrGlyAlaProGlyIleAlaGlyA 397
2233 GTCCGTCTCATCCGCGAGCGCAGACGCTCGTGTGCCGATCCATCGAAAGG 2184
397 laProGlyPheProGlyAla.....Arg 404
2183 AACGATTCGTGACCA...GGTTCGCGCCCGGCGCCCCCGCATGGTTCGAC 2137
405 GlyProSerGlyProGlnGlyProSerGlyAlaProGlyProLysGlyAs 421
2136 CTCGGTTCGCCCGATGTCGCCGCCTCGGCCGACTTCTACA..... 2097
421 nSerGlyGluProGlyAlaProGlyAsnLysGlyAspThrGlyAlaLysG 438
2096 .....CCGGCCTGTTCGGCTGGACCGCCA..... 2073
438 lyGluProGlyProAlaGlyValGlnGlyProProGlyProAlaGlyGlu 454
2072 .....CCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACGTTCAGC 2029
455 GluGlyLysArgGlyAlaArgGlyGluProGlyProAlaGlyLeuProGl 471
2028 TCCGACGGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCC 1979
471 yProAlaGlyGluArgGly.....AlaProGlySerArgGlyPheProG 486
978 CTACCACCGGCCGTACGGGCCCGGGAACGACCAGCACGGCATGCCGGCCA 1929
486 lyAlaAspGlyIleAlaGlyProLys.....GlyPro 496
1928 TCTGGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGG 1879
497 ProGlyGluArgGlySerProGlyAlaValGly.....ProLysGl 510
1878 GTCGAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGG 1829
510 ySerProGlyGluAlaGlyArgProGly.....GluAlaGlyLeuProG 525
1828 CCTCGGCCGGATGGCGGTCTTCGCCGACCCCGCCGGGGCCGCGTTCGCGG 1779
525 lyAlaLysGlyLeuThrGlySerProGlySerProGlyProAspGlyLys 541
1778 TCTGGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGC 1729
542 ThrGly.....ProPr 545
1728 TCGGTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCGGGA 1679
545 oGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProProGlyA 562
1678 CTTC.....TACCCGGCGACCCTCGGCCTGGCT 1651
562 laArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla 578
1650 .....CCGGCCGACACCGGACTGAAGGGCGTCACCGACCCGGTCTGGCA 1607
579 GlyGluProGlyLysProGlyGluArgGlyAlaProGlyProProGlyAl 595
1606 CATCGGTGACACACCGGTCGCCGGCACCCAG...GAGCTGGGCGTCACCG 1560
595 aValGly.....AlaAlaGlyLysAspGlyGluAlaGlyAlaGlnG 609
1559 GCGCGGTACGGCCGCACTGGGCCGTGCTGTTCGCCGTGCACGACTGCGAC 1510
609 lyProProGlyPro.ThrGlyPro.....AlaGl 618
1509 GCGACGG.....TCCGGCGCGCCCGTTGAACTCGGCGGCTC 1475
618 yGluArgGlyGluGlnGlyProAlaGlyAlaProGlyPheGlnGlyLeuP 635
1474 CGTCGAGAACGAGCCCGCCGACACGC..... 1449
635 roGlyProAlaGlyProProGlyGluAlaGlyLysProGlyGluGlnGly 651
1448 ...CCAGGGGGCGGCGGGCGGACCTGCTCGACCCGCACGGGGCCGGCTTC 1402
652 ValProGlyAsnAlaGly.....AlaProGlyProAlaGl 663
1401 TCGGTGGTCGAACTGCGGGAGGGGTACCCCGCGGCGGCGGGCGGTG.... 1356
663 y.....AlaArgGlyGluArgGlyPheProGlyGluArgGlyValGlnG 678
1355 .....CCTCGTGAGCGGCGAGGCGCCCCGGGTGGC 1326
678 lyProProGlyProGlnGlyProArgGlyAlaAsnGlyAlaProGly... 693
1325 CGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGCCCGAGG 1276
694 .....AsnAspGlyAlaLysGlyAspAlaGl 702
1275 TGCACGACGCCTTCCGGGA.....GGCGGGC 1250
702 yAlaProGlyAlaProGlyAsnGluGlyProProGlyLeuGluGlyMetP 719
1249 CCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATCAC 1200
719 roGlyGluArgGly.....AlaAlaGlyLeuProGlyAlaLysGlyAsp 733
1199 CGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAAGG 1150
734 ArgGlyAspProGlyProLysGlyAlaAspGlyAlaProGlyLysAspGl 750
1149 ACCCCGA.....CCTCGCCCCCGCCGCCTGG 1124
750 yLeuArgGlyLeuThrGlyProIleGlyProProGlyProAlaGlyAlaP 767
1123 CGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTGCGTCCGTT 1074
767 roGlyAsp..... 769
1073 CACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCG..... 1032
770 .....LysGlyGluAlaGlyProProGlyProAlaGlyProTh 782
1031 .....CCGCATCCACGCACCT 1016
782 rGlyAlaArgGlyAlaProGlyAspArgGlyGluProGlyPro..... 796
1015 GCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGAT 966
797 .....ProGlyProAlaGlyPheAla...GlyProProGlyAlaAsp 809
965 CGCCGGCCGGCTGCTCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCA 916
810 GlyGlnProGlyAlaLys.....GlyGl 817

```
 915 AACCGGCCGAGCTGATCGGCGG......CTTCGCGTACCACTTCCCGCTG 872
     :||||||   |||      :::                  |||   ||||||
 817 uThrGlyAspAlaGlyAlaLysGlyAspAlaGlyProProGlyProAla. 833

 871 TTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCCGGCGATGGC 822
       |||         |||         ||||||:::   ::::::   |||
 834 ..GlyProThrGlyAlaProGlyProAlaGly***ValGlyAlaProGly 849

 821 CCGCGAGGCCGTCAGCGTTCTCAAGGC..................... 795
     |||:::|||   :::   |||   |||
 850 ProLysGlyAlaArgGlySerAlaGlyProProGlyAlaThrGlyPhePr 866

 795 ..................................................... 795

 866 oGlyAlaAlaGlyArgValGlyProProGlyProSerGlyAsnIleGlyL 883

 794 ..ACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCC 747
           |||       ||||||   ::::::|||         ||||||
 883 euProGlyProProGlyProAlaGlyLys***Gly...SerLysGlyPro 898

 746 TGCCGGGGGCGTGCCGGACACCTCGGC...CCTGGAGAGCCTGCTCCTCG 700
       ||||||::::::|||      |||   ||||||||||||:::|||
 899 ...ArgGlyGluThrGlyProAlaGlyArgProGlyGluProGlyProAl 914

 699 AAGCCGTGCACTCAGCCCGGCGGAACGA...CACCCCGACCATGAC.... 657
     .  :::             |||:::|||::      |||      |||
 914 aGlyProProGlyProProGlyGluLysGlySerProGlyAlaAspGlyP 931

 656 .........................CCGCGTGCTGTACGAGCGCGCGCA 633
                              |||:::::::::   ::::::
 931 roIleGlyAlaProGlyThrProGlyProGlnGlyIleAlaGlyGlnArg 947

 632 GGCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCG 583
     |||   |||   |||             |||::::::
 948 GlyValValGlyLeu............ProGlyGln.............. 955

 582 GGCTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTG 533
                       |||   :::|||   :::|||||||||||||:
 956 ...................ArgGlyGluArgGlyPheProGlyLeuProG 966

 532 CTCGCGGAGGT...CCTGGC......GGGCCGCCTCGCGGCGGATGCCGA 492
     ::   :::|||   ||||||      ||||||      ::::::   :::
 966 lyProSerGlyGluProGlyLysGlnGlyProSerGlyAlaSerGlyGlu 982

 491 CGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGC 442
     ||||||         |||:::         |||   |||   ||||||::
 983 ArgGlyProProGlyProMetGlyProProGlyLeuAlaGlyProProGl 999

 441 CGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGC 392
     :   |||   :::                 ::::::|||:::      |||
 999 yGluAlaGlyArgGluGlyAlaProGlyAlaGluGlyAlaProGlyArgA 1016

 391 GGCGTCCGGCTGCCCCGCGGAGC..................... 369
           |||:::|||::::::
 016 spGlyAlaAlaGlyProLysGlyAspArgGlyGluThrGlyProAlaGly 1032

 368 ..............................................GC 367
                                                    ||
1033 ProProGlyAlaProGlyAlaProGlyAlaProGlyProValGlyProAl 1049

 366 CGGTGCTGGTGGACATCGAGGGCACCAA............CACCGACGGC 329
     ||||   :::|||   ||||||:::                      |
1049 aGlyLysAsnGlyAspArgGlyGluThrGlyProAlaGlyProAlaGlyP 1066

 328 CGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTC....CCTCGT 283
     |||||:::         ||||||   |||   ||||||      ||||||
1066 roProGlyProAlaGlyAlaArgGlyProAlaGlyProGlnGlyProArg 1082

 282 GGCGGC.......................................... 277
     |||
1083 GlyAspLysGlyGluThrGlyGluGlnGlyAspArgGlyMetLysGlyHi 1099

 276 ............................GGCTCACCTTCGGCGACGG 258
```

```
                                          |||   |||:::|||   |
1099 sArgGlyPheSerGlyLeuGlnGlyProProGlyProProGlyAlaProG 1116

 257 GCCGCACTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGA 208
     ||:::         ||||||:::|||:::|||:::      |||:::
1116 lyGluGlnGlyProSerGlyAlaSerGlyProAlaGlyProArgGlyPro 1132

 207 TGATCGGCGTACTGCGCAGCAGGTTCC.................CCGAG 176
        ::::::      |||||||||:::                  |||
1133 ProGlySerAlaGlyAlaAlaGlyLysAspGlyLeuAsnGlyLeuProGl 1149

 175 GCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAAGGGGGC 126
      |||      |||                                  ||||
1149 yProIleGlyPro.............................ProGlyP 1156

 125 CCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGCTGATCCGCCG 76
     |||||   |||   :::         |||   |||         ||||||
1156 roArgGlyArgThrGlyGluValGlyProValGlyProProGlyProPro 1172

  75 GCCGCGACCGGC...ACGGGACCGCCCACC 49
     :::      |||      |||||||||:::
1173 GlyProProGlyProProGlyProProSer 1182
```

```
seq_name: SwissProt_40:CA11_CANFA

seq_documentation_block:
ID   CA11_CANFA     STANDARD;      PRT;  1460 AA.
AC   Q9XSJ7;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RT   "Sequence of normal canine COL1A1 cDNA.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF153062; AAD34619.1; -.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   PROPEP       23    157       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       158   1214       COLLAGEN ALPHA 1(I) CHAIN.
```

```
FT   PROPEP      1215   1460       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN        34     92       VWFC.
FT   DOMAIN       158    174       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN       175   1188       TRIPLE-HELICAL REGION.
FT   DOMAIN      1189   1214       NONHELICAL REGION (C-TERMINAL).
FT   SITE         741    743       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE        1089   1091       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD    1361   1361       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1460 AA;  138762 MW;  58E3674D2B570697 CRC64;


alignment_scores:
            Quality:  464.50              Length:   1115
              Ratio:  0.940                Gaps:     61
[redacted]rcent Similarity:  44.305   Percent Identity: 26.547

[redacted]gment_block:
[redacted]09-673-254-1/rev x CA11_CANFA   ..

 Align seg 1/1  to: CA11_CANFA  from: 1  to: 1460

    2780 CCAGGT............CAATCACCTCGCAGGCTTCCTGCTGGCACGGC 2743
         ||||||           :::::::||||||   ||||||:::      ::
     174 ProGlyProMetGlyProSerGlyProArgGlyLeuProGlyProProGl 190

    2742 TGCTGCGGGA..............GCGGCTCGCGGGCGGGCGGCTGATC 2708
         :|||   |||               :::      |||:::   ::::::|
     190 yAlaProGlyProGlnGlyPheGlnGlyProProGlyGluProGlyGluP 207

    2707 CT............CACCTCGTCCGACGCGTACACCCAGGGCCG...... 2676
         ||            :::    |||      |||||||||
     207 roGlyAlaSerGlyProMetGlyProArgGlyProProGlyProProGly 223

    2675 .......................GATCGACCCGGACGACCTCAACGGCG 2650
                                :::|||||||||||||   :::||
     224 LysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProGlyGluAr 240

    2649 A......CCGTCACCGCTACAGCGCCGG......CCAGGCGTACGGCACG 2612
         .  |      |||:::|||         |||      ||||||:::      :
     240 gGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyL 257

    2611 TCCAAACAGGCCAACATCATGACCGCGGC............GGAGGCCGC 2574
         ::      |||:::       . ||||||              |||
     257 eu...ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAla 272

    2573 CAGGCGCTGGCCGGACGTGCTGGC........................... 2550
         ::::::   ||||||   ||||||
     273 LysGlyAspAlaGlyProAlaGlyProLysGlyGluProGlySerProGl 289

    2549 ..............GGTCAGCTATCACCCCGGTG.............A 2529
                       ||||||::::::||||||               |
     289 yGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGluA 306

    2528 GGTCCGCACCCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGT 2479
         ||:::   |||::::::   |||     :::   |||   ::::::|||
     306 rgGlyArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGly 322

    2478 TCAACCCCTTCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGG 2429
         :::|||   :::::::::|||            |||   |||:::
     323 AlaThrGlyAlaAlaGlyPro............ProGlyProThr..... 333

    2428 CTGGCGTCCGCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGA 2379
                |||      |||   :::   |||   |||:::   :::
     334 ....GlyProAlaGlyProProGlyPheProGlyAlaValGlyAlaLysG 349

    2378 CCGGCGGCTGTCCCCGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGG 2329
               :::   |||:::   |||                |||   :::
     349 lyGluAlaGlyProGlnGlyAlaArgGlySerGluGlyProGlnGlyVal 365

    2328 CGAAGCTCTGGGAGGCCGGCGCGGCCG..............CCGTCGGC 2294
         |||::::::      |||      |||               |||:::::
     366 ArgGlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGl 382

    2293 GACACCGCGCACTGACGGC......GGCGGCCCGCCCCGCCCGCATGTCC 2250
         ::::|||   :::||||||      ||| |||:::   |||:::   :::
     382 yAsnProGlyAlaAspGlyGlnProGly.AlaLysGlyAlaAsnGlyAla 398

    2249 GTCTCATCCGCGAGATGTCCGTCTCATCCGCGAGCGCAGACGCTCGTGTG 2200
            :::   |||   :::|||:::   |||   |||:::
     399 ProGlyIleAlaGlyAlaProGlyPheProGlyAlaArg..... 411

    2199 CCGATCCATCGAAAGGAACGATTCGTGACCAGGTTCGCGCCCGGCGCCCC 2150
                                                |||:::   |
     412 ................................GlyProSerGlyP 416

    2149 CGCATGGTTCGACCTCGGTTCGCCCGATGTCGCCGCCTCGGCCGACTTCT 2100
         ||:::|||               |||   :::||||||   |||   :::
     416 roGlnGly.............ProSerGlyProProGlyProLysGly 427

    2099 ACACCGGCCTGTTCGGCTGGACCGCCACCGTGGTCAGCGACCCGGGCGCC 2050
         :::   :::      :::::::|||         :::   |||   |||
     428 AsnSerGlyGluProGlyAlaProGlyAsnLysGlyAspThrGlyAlaLy 444

    2049 GGGGGATACACGACGTTCAGCTCCGACGGGAAGCTGGTCGCCGCGGTCGC 2000
          |||:::           :::|||||||||      :::|||   :::|
     444 sGlyGlu.........ProGlyProThrGlyIleGlnGlyProProGlyP 458

    1999 CCGCC.....................ACCAGATCGACACCCCCTACCACC 1971
         |||||                    :::|||:::   |||      |||
     458 roAlaGlyGluGluGlyLysArgGlyAlaArgGlyGluProGlyProThr 474

    1970 GGC.....................CGTACGGGCCCGGGAACG........ 1950
         |||                    |||   |||||||||:::
     475 GlyLeuProGlyProProGlyGluArgGlyGlyProGlySerArgGlyPh 491

    1949 ........ACCAGCACGGCATGCCGGCCATCTGGACCGTGTACTTCGCCAC 1907
                 ::::::   ::::::   |||:::||||||:::   :::
     491 eProGlyAlaAspGlyValAlaGlyProLysGlyProAlaGlyGluArgG 508

    1906 CGACGACGCCGACGCACTGACCAAGCGGGTCGAGACGGCGGGCGGCGAGG 1857
           :::               |||:::||||||      :::|||:::|||
     508 lySerProGlyProAlaGlyProLysGlySerProGlyGluAlaGlyArg 524

    1856 TCATCATGACTCCGATGGACGTCCTCGGCCTCGGCCGGATGGCGGTCTTC 1807
            :::      :::            ::::::   |||      :::||
     525 ProGly......GluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySe 539

    1806 GCCGACCCCGCCGGGGGCCGCGTTCGCGGTCTGGCGCAAGGGAGTCATGGA 1757
         ||||      |||||||||   ::::::::::|||
     539 rProGlySerProGlyProAspGlyLysThrGly..... 550

    1756 GGGCGCGGAGGTGACGGGCGTGCCCGGCTCGGTCGGCTGGGTCGAGCTGG 1707
                                    |||         ||||||::::::
     551 ......................ProProGlyProAlaGlyGlnAspGly 559

    1706 TGACCGACGGCATCGGGGCCGCCCGGGACTTCTACCCGGCGACCCTCGGC 1657
            |||          |||||||||||||:::          |||   :::||
     560 ArgProGlyProProGlyProProGlyAla.........ArgGlyGlnAl 573

    1656 CTGGCTCCGGCCGACACCGGACTGAAGGGCGTCACCGACCCGGTCTGGCA 1607
         |   :::          |||      :::::::::      :::   |||
     573 aGlyValMetGlyPheProGlyProLysGlyAlaAlaGlyGluProGlyL 590

    1606 CATCGGTGACACACCGGTCGCCGGCACCCAGGAGCTGGGCGTCACCGGCG 1557
           :::      :::      |||:::|||   :::   ::::::||||||
     590 ysAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyProAla 606

    1556 ............CGGTACGGCCGCACTGGGCCGTGCTGTTCGCCGTGCAC 1519
                     :::   |||      ||||||      :::|||:::
     607 GlyLysAspGlyGluAlaGlyAlaGlnGlyPro...ProGlyProAlaGl 622

    1518 GACTGCGACGCGACGG..............TCCGGCGCGCCGTTGAACT 1484
            |||   :::|||                :::|||:::|||      :
     622 yProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheG 639

    1483 CGGCGGCTCCGTCGAGAACGAGCCCGCCGACACGCCCAGGGGGCGGCGGG 1434
```

```
      :::::   |||:::   ::::::||||||   :::   |||      |||
  639 lnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProGly 655

 1433 CGG......ACCTGCTCGACCCGCACGGGGCCGGCTTCTCGGTGGTCGAA 1390
      :::      :::   :::         :::|||:::         :::::
  656 GluGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAlaAr 672

 1389 CTGCGGGAGGGGTACCCCGCGGCGGCGGGCGGTG..............  1356
      :      ||||||   |||   :::|||:::|||
  672 gGlyGluArgGlyPheProGlyGluArgGlyValGlnGlyProProGlyP 689

 1355 ........CCTCGTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTT 1314
              ||||||      :::||||||||||||:::
  689 roAlaGlyProArgGlyAlaAsnGlyAlaProGlyAsnAsp......... 702

 1313 CGCGTGTCCCATGATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCT 1264
        ::::::   |||      |||      |||   |||::::::
  703 GlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAl 719

 1263 TCCGGGA......GGCGGGCCCGGTCGTCGAGGTGAACGC..........  1230
       ||||||      |||   ||||||:::||||||
  719 aProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProG 736

 1229 ..CCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCG 1182
        |||:::|||                  |||      |||
  736 lyProLysGly...................AspArgGlyAspAlaGlyPro 746

 1181 CGAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGA........... 1143
      :::||||||      ||||||   :::::::|||   |||
  747 LysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyLeuThrGl 763

 1142 .............CCTCGCCCCCGCCGCCTGGCGGGGGGT...GGACGAC 1109
                  |||   |||         |||||||::   |||   :
  763 yProIleGlyProProGlyPro.........AlaGlyAlaProGlyAspL 777

 1108 GGTCTCGACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGC... 1062
      ::::::::         ||||||   |||:::      :::   |||
  777 ysGlyGluAlaGlyProSerGlyProAlaGlyProThrGlyAlaArgGly 793

 1061 ............CGTGGACGGCGAGGCCCACCGGCGCCTGCG........ 1032
                 ||||||:::   |||||||||:::||||||
  794 AlaProGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAl 810

 1031 ..................CCGCATCCACGCACCTGCGTTCAACCCGCGC 1001
                          :::|||   :::      :::|||:::
  810 aGlyProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyA 827

 1000 CGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGCT 951
        ||||||    :::   ||| |||:::|||   :::|||||||||
  827 spAlaGly.AlaLysGlyAsp.AlaGlyProProGlyProAlaGly.... 841

  950 CACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTGA 901
        |||::::::|||      |||   |||::::::::::   |||:::
  842 ..ProThrGlyProProGlyProIleGlyAsnValGlyAlaProGlyPro 857

  900 TCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCTC 851
      ::::::||| |||                  |||
  858 LysGlyAla.ArgGlySerAlaGlyProProGlyAlaThrGlyPheProG 874

  850 GGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTCT 801
        :::|||||:::   :::      ||||||   |||:::
  874 lyAlaAlaGlyArgValGlyProProGlyProSerGlyAsnAlaGlyPro 890

  800 CAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGG..... 756
        |||      |||      ||||||   ::::::|||   :::
  891 ProGly......ProProGlyProAlaGlyLysGluGlyGlyLysGlyAl 905

  755 ..........CACGGACCCTGCCGGGGGCGTGCCGGACACCTCGGC...C 719
                |||||||:::   ||||||   |||:::::|||   |
  905 aArgGlyGluThrGlyProAla...GlyArgProGlyGluValGlyProP 921

  718 CTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCCGGCGGAACGACAC 669
      |||||   |||:::|||   ::::::      |||||||:::
```

```
  921 roGlyProProGlyProAlaGlyGluLysGlySerProGlyAlaAspGly 937

  668 CCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCT 619
      |||         |||
  938 ProAlaGlyAlaPro..................................... 942

  618 CGGTCTCCGACGACCAGCTCGTCTACATGATCACCGGGCTCATCTTCGCC 569
       |||         |||      :::      :::|||:::   :::
  943 .GlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGlyL 959

  568 GGCCA......CGACACCACCGGCTCCTTCCTGGGCTTCCTGCT...... 531
        |||      |||   :::|||   :::||||||||||||:::
  959 euProGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSer 975

  530 ..........................................CGCGGAGG 523
                                                ||||||
  976 GlyGluProGlyLysGlnGlyProSerGlyThrSerGlyGluArgGlyPr 992

  522 TCCTGGC.......................GGGCCGCCT.......... 507
       ||||||                     |||||||||
  992 oProGlyProMetGlyProProGlyLeuAlaGlyProProGlyGluSerG 1009

  506 ..CGCGGCGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGA... 462
        |||   |||         :::::      |||   |||   |||
 1009 lyArgGluGlySerProGlyAlaGluGlySerProGlyArgAspGlySer 1025

  461 .....................................GGCGCTGCGCTACCACCCGC 442
                                          |||   |||   ||||||::
 1026 ProGlyProLysGlyAspArgGlyGluThrGlyProAlaGlyProProGl 1042

  441 C......GGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGTGACC 398
       :      ||||||
 1042 yAlaProGlyAla..................................... 1046

  397 ATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGA 348
                 |||:::||||||      :::|||   :::|||   :::
 1047 ...........ProGlyAlaProGlyProValGlyProAlaGlyLysAsn 1059

  347 GGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTCCACC 298
        |||   :::   :::      |||:::                  |||
 1060 GlyAspArgGlyGluThrGlyProAlaGly..............ProAl 1071

  297 CGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCACTAC 248
       ||||||         ::::::|||
 1071 aGlyProIleGlyProValGlyAla..................... 1079

  247 TGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCGGCGT 198
            ||||||   |||   |||   |||          |||      :::
 1080 .....ArgGlyProAlaGlyProGlnGlyProArgGlyAspLysGlyGlu 1094

  197 ACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGC...CGTGCCGTACGACG 151
        |||:::::|||      ||||||      |||   |||:::
 1095 ThrGlyGluGlnGlyAspArgGlyIleLysGlyHisArgGlyPheSerGl 1111

  150 AGTTGCGGTGGTGCCG.....................GAAGGGGGCCCAG 122
       :::         |||                    |||   ||||||:
 1111 yLeuGlnGlyProProGlyProProGlySerProGlyGluGlnGlyProS 1128

  121 ACGGCGCGGCTCA..........CCGAACTGCCCGTCTGGCTGCGCTGAT 82
      ::||||||:::            |||:::   |||   |||   |||
 1128 erGlyAlaSerGlyProAlaGlyProArgGlyProProGlySerAlaGly 1144

   81 CCGCCGGCCGCGACCGGC.....................ACGGGACCGCC 53
        |||:::      |||                      |||||||
 1145 SerProGlyLysAspGlyLeuAsnGlyLeuProGlyProIleGlyProPr 1161

   52 CACCGCGCACCGCGCGGTGGGC......GGTCCCGTG 22
      |      :::      :::|||      |||||||||
 1161 oGlyProArgGlyArgThrGlyAspAlaGlyProVal 1173

seq_name: SwissProt_40:CA11_HUMAN

seq_documentation_block:
```

```
ID   CA11_HUMAN     STANDARD;      PRT;  1464 AA.
AC   P02452; Q15176; Q14037;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-472 FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
     Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
     Jaenisch R., Prockup D.J.;
     "Structure of a full-length cDNA clone for the prepro alpha 1(I)
     chain of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [2]
RP   SEQUENCE OF 1-181 FROM N.A.
RX   MEDLINE=84270697; PubMed=6462220;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA   Myers J., Williams C., Ramirez F.;
RT   "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT   conservation of a pattern of introns and exons.";
RL   Nature 310:337-340(1984).
RN   [3]
RP   SEQUENCE OF 162-301.
RC   TISSUE=Skin;
RX   MEDLINE=71038625; PubMed=5529814;
RA   Click E.M., Bornstein P.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1 and alpha 2 chains of human skin collagen.";
RL   Biochemistry 9:4699-4706(1970).
RN   [4]
RP   SEQUENCE OF 263-268.
RC   TISSUE=Skin;
RX   MEDLINE=71001508; PubMed=4319110;
RA   Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT   "A comparative study of glycopeptides derived from selected
RT   vertebrate collagens. A possible role of the carbohydrate in fibril
RT   formation.";
RL   J. Biol. Chem. 245:5042-5048(1970).
RN   [5]
     SEQUENCE OF 425-1464 FROM N.A.
     MEDLINE=84080385; PubMed=6689127;
     Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
     Prockop D.J.;
     "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT   pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT   of structures that are conserved during evolution.";
RL   Biochemistry 22:5213-5223(1983).
RN   [6]
RP   SEQUENCE OF 1229-1454 FROM N.A.
RC   TISSUE=Bone;
RX   MEDLINE=88124208; PubMed=3340531;
RA   Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT   "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT   domain.";
RL   Nucleic Acids Res. 16:349-349(1988).
RN   [7]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=88097389; PubMed=3480516;
RA   Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT   "Regulatory elements in the first intron contribute to
RT   transcriptional control of the human alpha 1(I) collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN   [8]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=85130970; PubMed=2857713;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [9]
RP   SEQUENCE OF 1-44 FROM N.A.
RX   MEDLINE=88033098; PubMed=2822714;
RA   Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA   de Wet W.J.;
RT   "DNA sequences in the first intron of the human pro-alpha 1(I)
RT   collagen gene enhance transcription.";
RL   J. Biol. Chem. 262:15151-15157(1987).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=91184577; PubMed=2010058;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [12]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=91374476; PubMed=1895312;
RA   Byers P.H., Wallis G.A., Willing M.C.;
RT   "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL   J. Med. Genet. 28:433-442(1991).
RN   [13]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=97169389; PubMed=9016532;
RA   Dalgleish R.;
RT   "The human type I collagen mutation database.";
RL   Nucleic Acids Res. 25:181-187(1997).
RN   [14]
RP   VARIANT OI-II CYS-1166.
RX   MEDLINE=86287390; PubMed=3016737;
RA   Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT   "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT   change in one human pro alpha 1(I) collagen allele.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN   [15]
RP   VARIANT OI-II ARG-569.
RX   MEDLINE=87222295; PubMed=3108247;
RA   Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT   "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT   arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT   collagen.";
RL   J. Biol. Chem. 262:7021-7027(1987).
RN   [16]
RP   VARIANT OI-II CYS-926.
RX   MEDLINE=88033031; PubMed=3667599;
RA   Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT   "A point mutation in a type I procollagen gene converts glycine 748
RT   of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT   a lethal variant of osteogenesis imperfecta.";
RL   J. Biol. Chem. 262:14737-14744(1987).
RN   [17]
RP   VARIANT OI-II ARG-842.
RX   MEDLINE=88298828; PubMed=3403550;
RA   Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT   "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT   chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT   the peptide defect by in vitro expression of the mutant cDNA.";
RL   J. Biol. Chem. 263:11627-11630(1988).
RN   [18]
RP   VARIANT OI CYS-1195.
RX   MEDLINE=89218628; PubMed=3244312;
RA   Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT   "A cysteine for glycine substitution at position 1017 in an alpha
```

```
RT   1(I) chain of type I collagen in a patient with mild dominantly
RT   inherited osteogenesis imperfecta.";
RL   Mol. Biol. Med. 5:197-207(1988).
RN   [19]
RP   VARIANT OI-II VAL-434.
RX   MEDLINE=89255493; PubMed=2470760;
RA   Patterson E., Smiley E., Bonadio J.;
RT   "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT   mutation.";
RL   J. Biol. Chem. 264:10083-10087(1989).
RN   [20]
RP   VARIANT OI-IV SER-1010.
RX   MEDLINE=89308591; PubMed=2745420;
RA   Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT   "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT   one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL   J. Biol. Chem. 264:11893-11900(1989).
RN   [21]
RP   VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX   MEDLINE=89380165; PubMed=2777764;
     Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
     "Characterization of point mutations in the collagen COL1A1 and
     COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
     J. Biol. Chem. 264:15809-15812(1989).
RN   [22]
RP   VARIANT OI SER-1022.
RX   MEDLINE=90062068; PubMed=2511192;
RA   Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT   "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT   variant of osteogenesis imperfecta minimally destabilizes the triple
RT   helix of type I procollagen. The effects of glycine substitutions on
RT   thermal stability are either position of amino acid specific.";
RL   J. Biol. Chem. 264:19694-19699(1989).
RN   [23]
RP   VARIANT OI-II CYS-1082.
RX   MEDLINE=89109573; PubMed=2913053;
RA   Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT   "A lethal variant of osteogenesis imperfecta has a single base
RT   mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT   chain of type I procollagen. The asymptomatic mother has an
RT   unidentified mutation producing an overmodified and unstable type I
RT   procollagen.";
RL   J. Clin. Invest. 83:574-584(1989).
RN   [24]
RP   VARIANTS OI CYS-272; CYS-704 AND CYS-896.
RX   MEDLINE=90009313; PubMed=2794057;
RA   Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA   Weiss L., Graham J.M., Byers P.H.;
RT   "Osteogenesis imperfecta. The position of substitution for glycine by
RT   cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT   type I collagen determines the clinical phenotype.";
RL   J. Clin. Invest. 84:1206-1214(1989).
RN   [25]
RP   VARIANT OI-II CYS-422.

alignment_scores:
           Quality:  462.00              Length:   1122
             Ratio:  0.937                 Gaps:     62
 Percent Similarity:  43.939   Percent Identity: 26.560

alignment_block:
 US-09-673-254-1/rev x CA11_HUMAN   ..

 Align seg 1/1  to: CA11_HUMAN  from: 1  to: 1464

    2780 CCAGGT............CAATCACCTCGCAGGCTTCCTGCTGGCACGGC 2743
         !!!!!!           :::::!!!!!!   !!!!!!:::      ::
     178 ProGlyProMetGlyProSerGlyProArgGlyLeuProGlyProGl 194

    2742 TGCTGCGGGA..............GCGGCTCGCGGGCGGGCGGCTGATC 2708
         :!!!   !!!            :::      !!!:::   :::::::|
     194 yAlaProGlyProGlnGlyPheGlnGlyProProGlyGluProGlyGluP 211

    2707 CT............CACCTCGTCCGACGCGTACACCCAGGGCCG...... 2676
         ||              :::      !!!      !!!!!!!!
     211 roGlyAlaSerGlyProMetGlyProArgGlyProProGlyProProGly 227

    2675 ......................GATCGACCCGGACGACCTCAACGGCG 2650
                                :::!!!!!!!!!!!!   :::!!
     228 LysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProGlyGluAr 244

    2649 A......CCGTCACCGCTACAGCGCCGG......CCAGGCGTACGGCACG 2612
         |      !!!:::!!!         !!!      !!!!!!:::      :
     244 gGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyL 261

    2611 TCCAAACAGGCCAACATCATGACCGCGGC............GGAGGCCGC 2574
         ::      !!!:::         !!!!!!              !!!
     261 eu...ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAla 276

    2573 CAGGCGCTGGCCGGACGTGCTGGC.......................... 2550
         ::::::   !!!!!!   !!!!!!
     277 LysGlyAspAlaGlyProAlaGlyProLysGlyGluProGlySerProGl 293

    2549 ..............GGTCAGCTATCACCCCGGTG............A 2529
                       !!!!!!:::::::!!!!!!             !
     293 yGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGluA 310

    2528 GGTCCGCACCCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGT 2479
         !!:::   !!!::::::   !!!      :::   !!!   :::::!!!
     310 rgGlyArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGly 326

    2478 TCAACCCCTTCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGG 2429
         :::!!!   :::::::::!!!            !!!   !!!:::
     327 AlaThrGlyAlaAlaGlyPro............ProGlyProThr..... 337

    2428 CTGGCGTCCGCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGA 2379
              !!!      !!!   :::   !!!   !!!:::   :::
     338 ....GlyProAlaGlyProGlyProGlyPheProGlyAlaValGlyAlaLysG 353

    2378 CCGGCGGCTGTCCC................................... 2365
            :::   !!!
     353 lyGluAlaGlyProGlnGlyProArgGlySerGluGlyProGlnGlyVal 369

    2364 CGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCTGGGAG 2315
         !!!            !!!!!!   !!!!!!:::
     370 ArgGlyGluProGlyProProGlyProAlaGly.................. 380

    2314 GCCGGCGCGGCCGCCGTCGGCGACACCGCGCACTGACGGC......GGCG 2271
            !!!      !!!::::::::::!!!   :::!!!!!!      !!!
     381 ....AlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGly. 395

    2270 GCCCGCCCCGCCCGCATGTCCGTCTCATCCGCGAGATGTCCGTCTCATCC 2221
         !!!:::   !!!:::   :::   :::   !!!   :::!!!:::   !!
     396 AlaLysGlyAlaAsnGlyAlaProGlyIleAlaGlyAlaProGlyPhePr 412

    2220 GCGAGCGCAGACGCTCGTGTGCCGATCCATCGAAAGGAACGATTCGTGAC 2171
         |   !!!:::
     412 oGlyAlaArg..................................... 415

    2170 CAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCCCGATG 2121
                !!!:::   !!!:::!!!               !!!
     416 ........GlyProSerGlyProGlnGly..............ProGly 424

    2120 TCGCCGCCTCGGCCGACTTCTACACCGGCCTGTTCGGCTGGACCGCCACC 2071
         :::!!!!!!   !!!   :::            :::::!!!   !!!
     425 GlyProProGlyProLysGly............AsnSerGlyGluPro.. 436

    2070 GTGGTCAGCGACCCGGGCGCCGGGGGATACA.........CGACGTTCAG 2030
            :::!!!      :::   !!!!!!!!            :::   :
     437 ....GlyAlaProGlySerLysGlyAspThrGlyAlaLysGlyGluProG 452

    2029 CTCCGACGGGAAGCTGGTCGCCGCGGTCGCCCGCC.............. 1995
         ::!!!:::!!!      :::!!!   :::!!!!!!
     452 lyProValGlyValGlnGlyProProGlyProAlaGlyGluGluGlyLys 468

    1994 ......ACCAGATCGACACCCCCTACCACCGGC.................. 1968
```

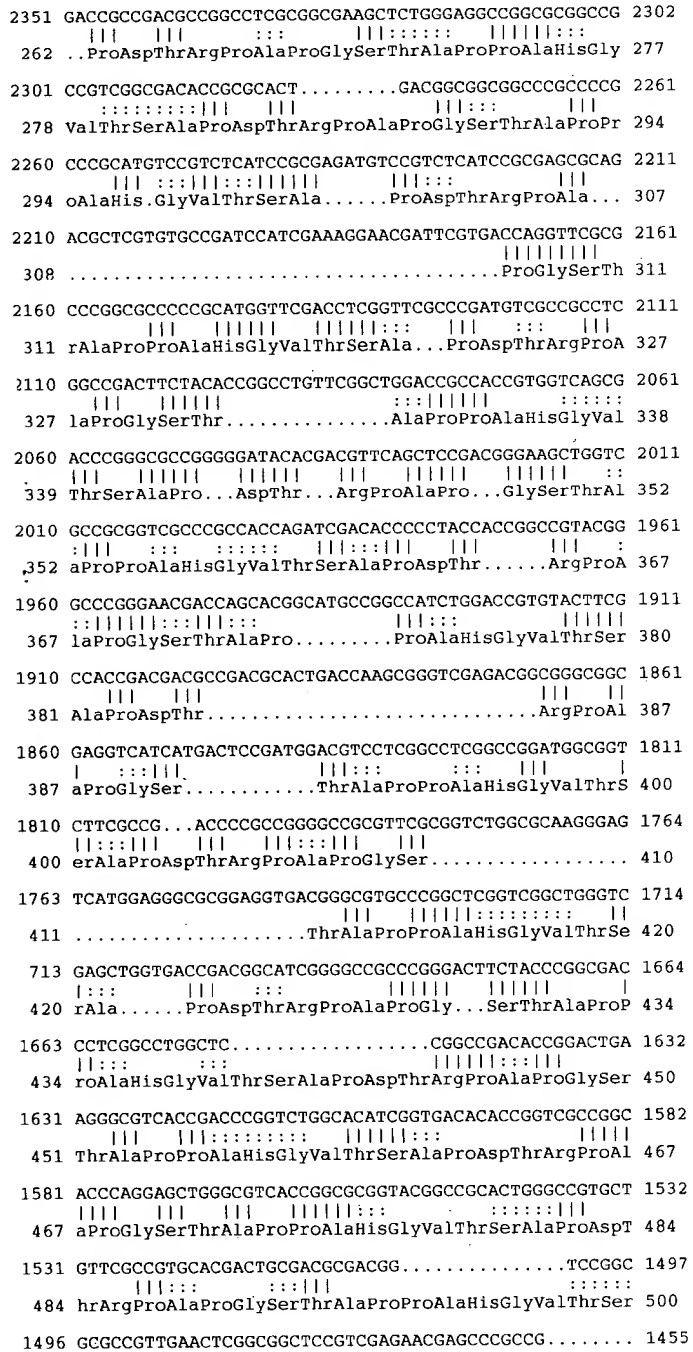
```
          :::|||:::   |||      ||||||
  469 ArgGlyAlaArgGlyGluProGlyProThrGlyLeuProGlyProProGl 485

 1967 ....CGTACGGGCCCGGGAACG..............ACCAGCACGGCAT 1937
          |||   |||||||||:::              ::::::   ::::
  485 yGluArgGlyGlyProGlySerArgGlyPheProGlyAlaAspGlyValA 502

 1936 GCCGGCCATCTGGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGA 1887
      ::   |||:::||||||:::   :::      :::
  502 laGlyProLysGlyProAlaGlyGluArgGlySerProGlyProAlaGly 518

 1886 CCAAGCGGGTCGAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGAC 1837
      |||:::||||||      :::|||:::|||   :::      :::
  519 ProLysGlySerProGlyGluAlaGlyArgProGly......GluAlaGl 533

  836 GTCCTCGGCCTCGGCCGGATGGCGGTCTTCGCCGACCCCGCCGGGGCCGC 1787
          ::::::   |||      :::||||||      |||||||||
  533 yLeuProGlyAlaLysGlyLeuThrGlySerProGlySerProGlyProA 550

 1786 GTTCGCGGTCTGGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACGGGCG 1737
        :::::::::|||
  550 spGlyLysThrGly..................................... 554

 1736 TGCCCGGCTCGGTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCC 1687
        |||         ||||||::::::      |||         |||||
  555 ...ProProGlyProAlaGlyGlnAspGlyArgProGlyProProGlyPr 570

 1686 GCCCGGGACTTCTACCCGGCGACCCTCGGCCTGGCTCCGGCCGACACCGG 1637
      |||||||:::          |||   :::|||   :::          |||
  570 oProGlyAla.........ArgGlyGlnAlaGlyValMetGlyPheProG 584

 1636 ACTGAAGGGCGTCACCGACCCGGTCTGGCACATCGGTGACACACCGGTCG 1587
          :::::::::      :::   |||   :::      :::
  584 lyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyGluArgGlyVal 600

 1586 CCGGCACCCAGGAGCTGGGCGTCACCGGCG............CGGTACGG 1549
      |||:::|||   :::   ::::::||||||           :::   ||
  601 ProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAlaGl 617

 1548 CCGCACTGGGCCGTGCTGTTCGCCGTGCACGACTGCGACGCGACGG.... 1503
      |      ||||||      :::|||:::      |||   :::|||
  617 yAlaGlnGlyPro...ProGlyProAlaGlyProAlaGlyGluArgGlyG 633

 1502 ..........TCCGGCGCGCCGTTGAACTCGGCGGCTCCGTCGAGAACG 1464
              :::|||:::|||      ::::::   |||:::   :::
  633 luGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAla 649

  463 AGCCCGCCGACACGCCCAGGGGGCGGCGGGCGG......ACCTGCTCGAC 1420
      :::||||||   :::   |||      |||:::      :::   :::
  650 GlyProProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAs 666

 1419 CCGCACGGGGCCGGCTTCTCGGTGGTCGAACTGCGGGAGGGGTACCCCGC 1370
          :::|||:::         ::::::      ||||||   |||
  666 pLeuGlyAlaProGlyProSerGlyAlaArgGlyGluArgGlyPheProG 683

 1369 GGCGGCGGGCGGTG.....................CCTCGTGAGCGG 1344
        :::|||:::|||                     ||||||
  683 lyGluArgGlyValGlnGlyProProGlyProAlaGlyProArgGlyAla 699

 1343 CGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCA 1294
      :::||||||||||||:::              ::::::   |||
  700 AsnGlyAlaProGlyAsnAsp.........GlyAlaLysGlyAspAlaGl 713

 1293 TGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGA......GGCGGGC 1250
        |||      |||   |||:::::      ||||||      |||   |
  713 yAlaProGlyAlaProGlySerGlnGlyAlaProGlyLeuGlnGlyMetP 730

 1249 CCGGTCGTCGAGGTGAACGC............CCCCGCGGGCGGACCCGC 1212
      |||||:::||||||                  |||:::|||
  730 roGlyGluArgGlyAlaAlaGlyLeuProGlyProLysGly......... 743

 1211 CTGGGTCATCACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCC 1162
                |||            |||:::||||||      |||||
```

```
  744 .........AspArgGlyAspAlaGlyProLysGlyAlaAspGlySerPr 757

 1161 GGTTCGTGAAGGACCCCGA.....................CCTCGCC 1136
      |   :::::::|||   |||                        |||   |
  757 oGlyLysAspGlyValArgGlyLeuThrGlyProIleGlyProProGlyP 774

 1135 CCCGCCGCCTGGCGGGGGGT...GGACGACGGTCTCGACATCCCCGTTCC 1089
      ||         ||||||:::   |||   :::::::::         |||
  774 ro.........AlaGlyAlaProGlyAspLysGlyGluSerGlyProSer 787

 1088 GGAGCTGCGTCCGTTCACGCTCATCGC.............CGTGGACG 1054
      |||   |||:::    :::   |||                ||||||::
  788 GlyProAlaGlyProThrGlyAlaArgGlyAlaProGlyAspArgGlyGl 804

 1053 GCGAGGCCCACCGGCGCCTGCG.......................C 1031
      :   |||||||||:::||||||
  804 uProGlyProProGlyProAlaGlyPheAlaGlyProProGlyAlaAspG 821

 1030 CGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGGA 981
        :::|||   :::      :::|||:::   ||||||    :::   ||
  821 lyGlnProGlyAlaLysGlyGluProGlyAspAlaGly.AlaLysGlyAs 837

  980 TCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCCT 931
      | |||:::|||   :::||||||||||      |||   :::|||
  837 p.AlaGlyProProGlyProAlaGly......ProAlaGlyProProGly 851

  930 CCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCAC 881
      |||   |||:::::::::   |||:::   ::::::||| |||
  852 ProIleGlyAsnValGlyAlaProGlyAlaLysGlyAla.ArgGlySerA 868

  880 TTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCC 831
                |||                 :::||||||:::   :::
  868 laGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArgValGly 884

  830 GGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGCG 781
          ||||||   |||:::          |||      |||
  885 ProProGlyProSerGlyAsnAlaGlyProProGly......ProProGl 899

  780 GCCCGCAGAGCGGCGGGGGTGACGG..............CACGGACCCT 746
       ||||||   :::::::|||   :::                ||||||:
  899 yProAlaGlyLysGluGlyGlyLysGlyProArgGlyGluThrGlyProA 916

  745 GCCGGGGGCGTGCCGGACACCTCGGC...CCTGGAGAGCCTGCTCCTCGA 699
      ::   ||||||   |||:::::|||   ||||||   |||:::|||
  916 la...GlyArgProGlyGluValGlyProProGlyProProGlyProAla 931

  698 AGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCGTGC 649
      ::::::      ||||||:::          |||         |||
  932 GlyGluLysGlySerProGlyAlaAspGlyProAlaGlyAlaPro..... 946

  648 TGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAGCTC 599
                                     |||         |||
  947 ............................GlyThrProGlyProGlnG 953

  598 GTCTACATGATCACCGGGCTCATCTTCGCCGGCCA......CGACACCAC 555
        :::      :::|||:::   :::      |||      |||   :::
  953 lyIleAlaGlyGlnArgGlyValValGlyLeuProGlyGlnArgGlyGlu 969

  554 CGGCTCCTTCCTGGGCTTCCTGCTCGCGGAGGT...CCTGGC......GG 514
      |||   :::||||||||||||:::   :::|||   ||||||      ||
  970 ArgGlyPheProGlyLeuProGlyProSerGlyGluProGlyLysGlnGl 986

  513 GCCGCCTCGCGGCGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAG 464
      ||||      ::::::   :::||||||          |||:::
  986 yProSerGlyAlaSerGlyGluArgGlyProProGlyProMetGlyProP 1003

  463 GAGGCGCTGCGCTACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGC 414
        |||   |||   ||||||
 1003 roGlyLeuAlaGlyProPro............................. 1009

  413 TGCCACGGAGGTGACCATCGGCGGCGT...................... 387
            |||:::      |||:::
 1010 .........GlyGluSerGlyArgGluGlyAlaProGlyAlaGluGlySe 1023
```

```
 386 .CCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAAC 338
     |||:::         |||   ||||||   |||   ||||||
1023 rProGlyArgAspGlySerProGlyAlaLysGlyAspArgGly..... 1037

 337 ACCGACGGCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCC 288
                 :::      ||||||      |||   :::|||:::
1038 .............GluThrGlyProAlaGlyProProGlyAlaProGly 1049

 287 CTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGG 238
           :::|||                                  ||
1050 AlaProGlyAlaPro..............................Gl 1055

 237 AGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGC 188
     |   :::   |||||||||   :::   |||      :::|||:::
1055 yProValGlyProAlaGlyLysSerGlyAspArgGlyGluThrGlyProA 1072

 187 AGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTG 138
          |||   ||||||:::|||   |||      |||      :::
1072 laGlyProAlaGlyProValGlyProAlaGlyAlaArgGlyProAlaGly 1088

 137 CCGGAAGGG..................................... 129
     |||:::|||
1089 ProGlnGlyProArgGlyAspLysGlyGluThrGlyGluGlnGlyAspAr 1105

 128 .....................................GGCCCAGACGGCG 117
                                          ||||||   |||
1105 gGlyIleLysGlyHisArgGlyPheSerGlyLeuGlnGlyProProGlyP 1122

 116 C....GGCTCACCG......AACTGCCCGTCTGGCTGCGCTGATCCGCCG 76
          |||||||||      :::   |||||||||::::::   |||
1122 roProGlySerProGlyGluGlnGlyProSerGlyAlaSerGlyPro... 1137

  75 GCCGCGACCGGCACGGGACCGCCCACCGCGCACCGCGCG........... 37
        |||        |||||||||   :::      |||
1138 ...AlaGlyProArgGlyProProGlySerAlaGlyAlaProGlyLysAs 1153

  36 .......................GTGGGCGGTCCCGTGCCGTGTACCC 12
                            :::|||   |||   |||      |
1153 pGlyLeuAsnGlyLeuProGlyProIleGlyProProGlyProArgGlyA 1170

  11 GGTGCGGA 4
     ||   |||
1170 rgThrGly 1172
```

```
seq_name: SwissProt_40:MUC1_HUMAN

seq_documentation_block:
ID   MUC1_HUMAN     STANDARD;      PRT;  1255 AA.
AC   P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT   01-JAN-1990 (Rel. 13, Created)
     01-APR-1990 (Rel. 14, Last sequence update)
     16-OCT-2001 (Rel. 40, Last annotation update)
     Mucin 1 precursor (Polymorphic epithelial mucin) (PEM) (PEMT)
     (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE   (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE   reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE   DF3).
GN   MUC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=90368716; PubMed=2394722;
RA   Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT   "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL   J. Biol. Chem. 265:15294-15299(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90202794; PubMed=2318825;
RA   Ligtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT   "Episialin, a carcinoma-associated mucin, is generated by a
RT   polymorphic gene encoding splice variants with alternative amino
RT   termini.";
RL   J. Biol. Chem. 265:5573-5578(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90368715; PubMed=1697589;
RA   Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA   Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT   "Molecular cloning and expression of human tumor-associated
RT   polymorphic epithelial mucin.";
RL   J. Biol. Chem. 265:15286-15293(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91097524; PubMed=2268309;
RA   Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA   Taylor-Papadimitriou J., Gendler S.J.;
RT   "Structure and expression of the human polymorphic epithelial mucin
RT   gene: an expressed VNTR unit.";
RL   Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90276413; PubMed=2351132;
RA   Wreschener D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA   Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA   Keydar I.;
RT   "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT   may generate multiple protein forms.";
RL   Eur. J. Biochem. 189:463-473(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90276414; PubMed=2112460;
RA   Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA   Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT   "A transcribed gene, containing a variable number of tandem repeats,
RT   codes for a human epithelial tumor antigen. cDNA cloning, expression
RT   of the transfected gene and over-expression in breast cancer
RT   tissue.";
RL   Eur. J. Biochem. 189:475-486(1990).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91033045; PubMed=1688329;
RA   Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA   Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT   "Isolation and characterization of an expressed hypervariable gene
RT   coding for a breast-cancer-associated antigen.";
RL   Gene 93:313-318(1990).
RN   [8]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=88330762; PubMed=3417635;
RA   Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA   Burchell J.;
RT   "A highly immunogenic region of a human polymorphic epithelial mucin
RT   expressed by carcinomas is made up of tandem repeats.";
RL   J. Biol. Chem. 263:12820-12823(1988).
RN   [9]
RP   SEQUENCE OF 1-169 FROM N.A.
RX   MEDLINE=90088473; PubMed=2597151;
RA   Abe M., Siddiqui J., Kufe D.;
RT   "Sequence analysis of the 5' region of the human DF3 breast
RT   carcinoma-associated antigen gene.";
RL   Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN   [10]
RP   SEQUENCE OF 1-109 FROM N.A.
RC   TISSUE=Thyroid;
RX   MEDLINE=96183746; PubMed=8608966;
RA   Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT   "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT   transcriptase polymerase chain reaction of the MUC1 gene.";
```

```
RL   Int. J. Cancer 66:55-59(1996).
RN   [11]
RP   SEQUENCE OF 1-89 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=96181716; PubMed=8604237;
RA   Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA   Lee L.N., Luh K.T., Wu C.W.;
RT   "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT   tissues.";
RL   Oncology 53:118-126(1996).
RN   [12]
RP   SEQUENCE OF 1-46 FROM N.A.
RC   TISSUE=Breast carcinoma;
RA   Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RA   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC       IS ALSO PRODUCED.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
CC       TUMORS, SUCH AS BREAST CANCER.
CC   -!- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC
CC       ACID).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
CC       VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
CC       MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05582; AAA60019.1; -.
DR   EMBL; M32738; AAA35804.1; -.
DR   EMBL; M32739; AAA35806.1; -.
DR   EMBL; J05581; AAA59876.1; -.
DR   EMBL; M61170; AAB53150.1; -.
DR   EMBL; X52229; CAA36478.1; ALT_SEQ.
DR   EMBL; X52228; CAA36477.1; ALT_SEQ.
     EMBL; M35093; AAB59612.1; ALT_SEQ.
     EMBL; Z17324; CAA78972.1; -.
     EMBL; Z17325; CAA78973.1; -.
     EMBL; M31823; AAA35757.1; -.
     EMBL; S81781; AAD14376.1; ALT_INIT.
DR   EMBL; S81736; AAD14369.1; ALT_INIT.
DR   EMBL; M21868; AAA59874.1; ALT_SEQ.
DR   PIR; A35175; A35175.
DR   PIR; B35175; B35175.
DR   PIR; S10218; S10218.
DR   GlycoSuiteDB; P15941; -.
DR   MIM; 158340; -.
DR   MIM; 113720; -.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat; Alternative splicing.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1255       MUCIN 1.
FT   DOMAIN       24   1162       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1163   1186       POTENTIAL.
FT   DOMAIN     1187   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       81    960       44 X 20 AA TANDEM REPEATS.
FT   DOMAIN     1034   1151       SEA.
FT   CARBOHYD    957    957       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    975    975       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1029   1029       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1055   1055       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1133   1133       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     19     19       T -> TATTAPKPAT (IN ISOFORM B).
FT   VARSPLIC     20     22       MISSING (IN ISOFORM C).
FT   VARSPLIC     20     31       MISSING (IN ISOFORM D).
FT   VARSPLIC    126    905       MISSING (IN POLYMORPHIC EPITHELIAL
FT                                ISOFORM).
FT   VARSPLIC   1077   1087       FLQIYKQGGFL -> VSIGLSFPMLP (IN SECRETED
FT                                ISOFORM).
FT   VARSPLIC   1088   1255       MISSING (IN SECRETED ISOFORM).
FT   CONFLICT      2      2       T -> A (IN REF. 11).
FT   CONFLICT    134    134       P -> Q (IN REF. 9).
FT   CONFLICT    154    154       P -> Q (IN REF. 9).
FT   CONFLICT   1021   1021       S -> T (IN REF. 3).
FT   CONFLICT   1251   1251       A -> T (IN REF. 3).
SQ   SEQUENCE   1255 AA;  122072 MW;  5E28DFC4DE7D9A82 CRC64;


alignment_scores:
           Quality:  455.50            Length:   1030
             Ratio:  0.967               Gaps:     64
 Percent Similarity:  45.728   Percent Identity: 28.738

alignment_block:
 US-09-673-254-1/rev x MUC1_HUMAN   ..

 Align seg 1/1  to: MUC1_HUMAN  from: 1  to: 1255

    2865 ACCCGCGCATCGATGTCATGGCCAGCAACGCCGGCGGCATGTTCTGGTCG 2816
         ||||||   :::::::::       :::   ||||||
     104 ThrArgProAlaLeuGlySerThrThrProProAlaHisAspValThrSe 120

    2815 CGCACCA......CGACCCAGGACGGGTTCG...AGGCCACCATCCAGGT 2775
          ||||||      :::|||      ||||||      ||||||:::::::
     120 rAlaProAspAsnLysProAlaProGlySerThrAlaProProAlaHisG 137

    2774 CAATCACCTCGCAGGCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCG 2725
         :::::||||||      :::         :::      :::::::::
     137 lyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaPro 153

    2724 CGGGCGGGCGGCTGATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGG 2675
            |||   |||   :::|||         ||||||   |||      ||
     154 ProAlaHisGlyValThrSerAlaProAspThrArg...ProAlaProGl 169

    2674 ATCGACCCGGACGACCTCAACGGCGACCGTCACCGCT......ACAGCGC 2631
         |||||||          :::   ::::::::::|||      |||   |
     169 ySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgP 186

    2630 CGGCCAGGCGTACGGCACGTCCA......AACAGGCCAACATCATGACCG 2587
         |||||      ::::::   |||      :::      ||||||   |||
     186 roAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaPro 202

    2586 CGGCGGAGGCCGCCAGGCGCTGGCCGGACGTGCTGGCGGTCAGCTATCAC 2537
              ||||||                                     ::
     203 AspThrArgPro.....................................Al 207

    2536 CCCGGTGAGGTCCGCACCCGCATCGGACGGGGCACGGTCG.........C 2496
         :|||      :::||||||   :::   |||:::   |||
     207 aProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspT 224

    2495 CTCGTCCTACTTCCGGTTCAACCCCTTCCTGCGCTCCGCGGCGAAGGGCG 2446
           ||||||:::   |||||||||         ||||||      ::::::
     224 hrArgProAlaProGlySerThr.........AlaProProAlaHisGly 237

    2445 CCGACACCCTCGTGTGGCTGGCGTCCGCGCCGGCCGAGGAGTTGACCACG 2396
            |||   :::         ||||||            :::   |||
     238 ValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPr 254

    2395 GGCG......GCTACTACAGCGACCGGCGGCTGTCCCCGGTGAGCGGCCC 2352
          |||      :::|||:::|||
     254 oAlaHisGlyValThrSerAla........................... 261
```


2351 GACCGCCGACGCCGGCCTCGCGGCGAAGCTCTGGGAGGCCGGCGCGGCCG 2302
262 ..ProAspThrArgProAlaProGlySerThrAlaProProAlaHisGly 277
2301 CCGTCGGCGACACCGCGCACT.....GACGGCGGCGGCCCGCCCCG 2261
278 ValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPr 294
2260 CCCGCATGTCCGTCTCATCCGCGAGATGTCCGTCTCATCCGCGAGCGCAG 2211
294 oAlaHis.GlyValThrSerAla.....ProAspThrArgProAla... 307
2210 ACGCTCGTGTGCCGATCCATCGAAAGGAACGATTCGTGACCAGGTTCGCG 2161
308 .....ProGlySerTh 311
2160 CCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCCCGATGTCGCCGCCTC 2111
311 rAlaProProAlaHisGlyValThrSerAla...ProAspThrArgProA 327
2110 GGCCGACTTCTACACCGGCCTGTTCGGCTGGACCGCCACCGTGGTCAGCG 2061
327 laProGlySerThr.....AlaProProAlaHisGlyVal 338
2060 ACCCGGGCGCCGGGGGATACACGACGTTCAGCTCCGACGGGAAGCTGGTC 2011
339 ThrSerAlaPro...AspThr...ArgProAlaPro...GlySerThrAl 352
2010 GCCGCGGTCGCCCGCCACCAGATCGACACCCCCTACCACCGGCCGTACGG 1961
352 aProProAlaHisGlyValThrSerAlaProAspThr.....ArgProA 367
1960 GCCCGGGAACGACCAGCACGGCATGCCGGCCATCTGGACCGTGTACTTCG 1911
367 laProGlySerThrAlaPro.....ProAlaHisGlyValThrSer 380
1910 CCACCGACGACGCCGACGCACTGACCAAGCGGGTCGAGACGGCGGGCGGC 1861
381 AlaProAspThr.....ArgProAl 387
1860 GAGGTCATCATGACTCCGATGGACGTCCTCGGCCTCGGCCGGATGGCGGT 1811
387 aProGlySer.....ThrAlaProProAlaHisGlyValThrS 400
1810 CTTCGCCG...ACCCCGCCGGGGCCGCGTTCGCGGTCTGGCGCAAGGGAG 1764
400 erAlaProAspThrArgProAlaProGlySer..... 410
1763 TCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGCTCGGTCGGCTGGGTC 1714
411 .....ThrAlaProProAlaHisGlyValThrSe 420
713 GAGCTGGTGACCGACGGCATCGGGGCCGCCCGGGACTTCTACCCGGCGAC 1664
420 rAla.....ProAspThrArgProAlaProGly...SerThrAlaProP 434
1663 CCTCGGCCTGGCTC.....CGGCCGACACCGGACTGA 1632
434 roAlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySer 450
1631 AGGGCGTCACCGACCCGGTCTGGCACATCGGTGACACACCGGTCGCCGGC 1582
451 ThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAl 467
1581 ACCCAGGAGCTGGGCGTCACCGGCGCGGTACGGCCGCACTGGGCCGTGCT 1532
467 aProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspT 484
1531 GTTCGCCGTGCACGACTGCGACGCGACGG.....TCCGGC 1497
484 hrArgProAlaProGlySerThrAlaProProAlaHisGlyValThrSer 500
1496 GCGCCGTTGAACTCGGCGGCTCCGTCGAGAACGAGCCCGCCG..... 1455
501 AlaProAspThrArgProAlaProGlySerThrAlaProProAlaHisGl 517
1454 .....ACACGCCCAGGGGGCGGC..... 1437
517 yValThrSerAlaProAspThrArgProAlaProGlySerThrAlaProP 534
1436 .....GGGCGGACCTGCTCGACCCGCACGGGGCCGGCTTCTCGGTGG 1395
534 roAlaHisGlyValThrSerAlaProAspThrArgProAlaPro..... 548
1394 TCGAACTGCGGGAGGGGTACCCCGCGGCGGCGGGCGGTGCCTCGT..... 1350
549 .....GlySerThrAlaProProAlaHisGlyValThrSerAlaPr 562
1349 .GAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCATG 1301
562 oAspThrArgProAlaProGlySerThrAlaPro..... 573
1300 ATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGGG 1251
574 .....ProAlaHisGlyValThrSerAlaProAspThrArg 585
1250 CCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATCA 1201
586 ProAlaProGlySer..ThrAlaProProAlaHisGlyValThrSerAla 601
1200 CCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAAG 1151
602 ProAspThrArg.....ProAlaProGlySer..... 610
1150 GACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCGA 1101
611 .....ThrAlaProProAlaHisGly.....ValThrS 620
1100 CATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGCG 1051
620 erAlaPro.....AspThrArgProAlaPro.GlySerTh 631
1050 AGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCGC 1001
631 rAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaP 648
1000 CGGCTGGCCGAGCGGACGGATCGCAT.....CGCCGCGATCGC 963
648 roGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThr 664
962 CGGCCGGCTGCTCACCGAACTCGCCGACGCCTCCGGCC...GGTCGGGCA 916
665 ArgPro...Ala..ProGlySerThrAlaProProAlaHisGlyValThr 679
915 AACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTC 866
680 SerAlaProAspThrArgProAlaProGlySerThr.....Al 692
865 ATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCG. 817
692 aProProAlaHisGlyValThrSerAlaProAspThrArg...ProAlaP 708
816 .....AGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGCGGCC 778
708 roGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThr 724
777 CGCAGAGCGGCGGGGGTGACGGCA..... 754
725 ArgProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAl 741
753 .....CGGACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGA 714
741 aProAspThrArgProAlaProGlySer.....ThrAlaProProAlaH 756
713 GAGCCTGCTCCTCGAAGCCGTGCACTCAG.....CCCGGCGGAACGACA 670

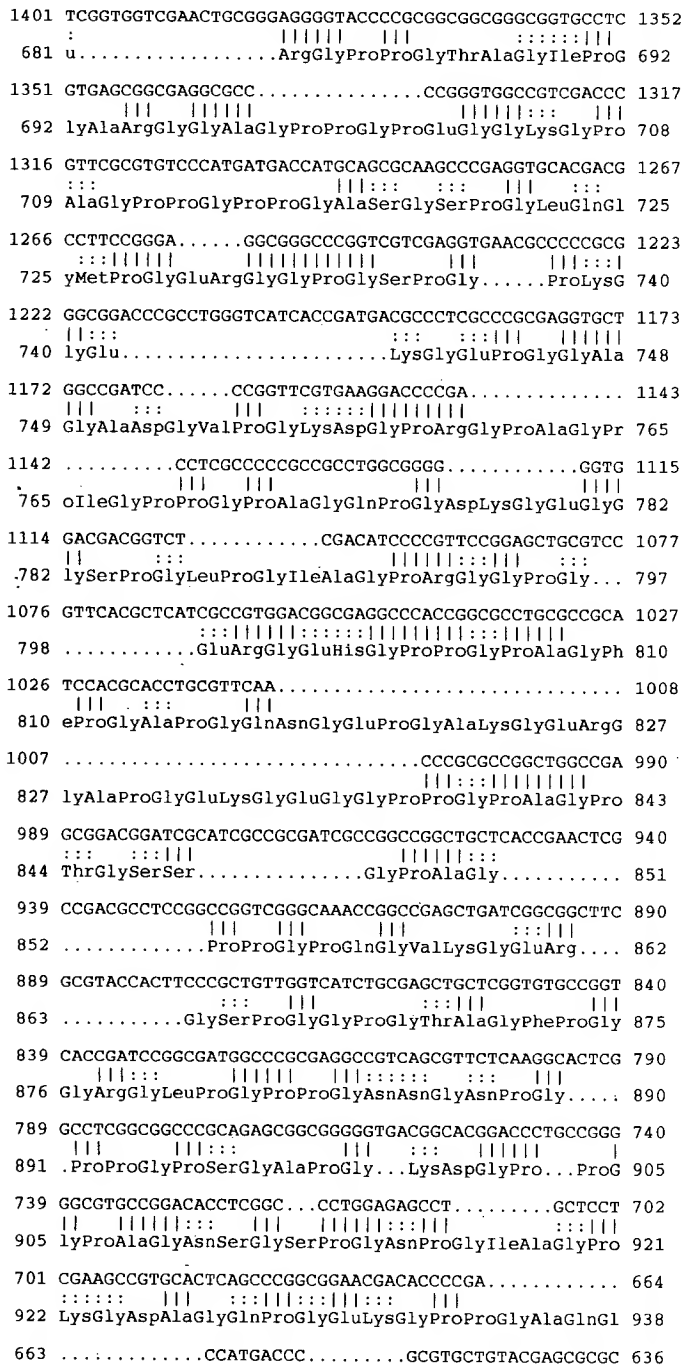
```
     756 isGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAla 772

     669 CCCCGACCATGACCCGCG...TGCTGTACGAGCGCGCGCAGGCCGAGTTC 623
         |||         ||||||          |||||||||       |||:::::
     773 Pro.........ProAlaHisGlyValThrSerAla......ProAspTh 784

     622 GGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCG......GGCT 579
         :      :::|||         :::||||||      |||      |||
     784 rArgProAlaPro.........GlySerThrAlaProProAlaHisGlyV 798

     578 CATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGCTCG 529
           :::|||         |||   |||||||||:::   ::::::
     798 alThrSerAlaProAspThrArgProAlaProGlySerThrAlaProPro 814

     528 CGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCCGTC 479
           ::::::      :::|||           |||         |||::
     815 AlaHisGlyValThrSerAla............ProAspThrArgProAl 827

     478 TCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTACAC 429
         :|||||||||                                |||::::
     827 aProGlySer.......................ThrAlaProProAlaH 836

     428 GTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGCTGC 379
         ::         |||   |||         |||:::   :::|||   :::
     836 isGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAla 852

     378 CCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGACGGC 329
         |||                                    |||:::
     853 Pro.....................................ProAlaHi 856

     328 CGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTGGCG 279
          ::::::|||:::      :::|||:::||||||   :::|||   |||:
     856 sGlyValThrSer......AlaProAspThrArgProAlaPro...GlyS 870

     278 GCGGCTCACCTTCGGCGACGGGCCGCACTACT..................... 247
         ::   :::|||   |||   |||   |||:::
     870 erThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgPro 886

     246 GCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCGGCGTA 197
         |||   ||||||::::::|||                       |||::
     887 AlaProGlySerThrAlaPro.....................ProAlaHi 896

     196 CTGCGCAGCAGGTTCCCCG......AGGCCCGACTGGCCGTGCCGTACGA 153
         :   :::::::::::::|||      ||||||               :::
     896 sGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAlaP 913

     152 CGAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAAC 103
               :::   |||:::      :::|||      |||   :::|||
     913 roProAlaHisGlyValThrSerAlaProAspThrArgProAlaPro... 928

     102 TGCCCGTCTGGCTGCGCTGATCCGCCGGCCGCGACCGGCACGGGACCGCC 53
                 |||   :::   |||||||||         |||:::   ||
     929 .........GlySerThrAlaProProAlaHisGlyValThrSerAlaPr 942

      52 CACCGCGCACCGCGCGGTGGGCGGTCCCGTGCCG 19
         |      :::   |||:::|||:::   :::|||
     942 oAspAsnArgProAlaLeuGlySerThrAlaPro 953

seq_name: SwissProt_40:CA11_MOUSE

seq_documentation_block:
ID   CA11_MOUSE     STANDARD;      PRT;  1453 AA.
AC   P11087; Q60635;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1 OR COLA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE=96033240; PubMed=8535610;
RA   Li S.W., Khillan J., Prockop D.J.;
RT   "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT   of type I procollagen.";
RL   Matrix Biol. 14:593-595(1995).
RN   [2]
RP   SEQUENCE OF 518-1128 FROM N.A.
RX   MEDLINE=86137403; PubMed=3841523;
RA   French B.T., Lee W.-H., Maul G.G.;
RT   "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT   collagen protein.";
RL   Gene 39:311-312(1985).
RN   [3]
RP   SEQUENCE OF 735-1130 FROM N.A.
RX   MEDLINE=83141374; PubMed=6298597;
RA   Monson J.M., Friedman J., McCarthy B.J.;
RT   "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT   evidence for a mouse B1 element within the gene.";
RL   Mol. Cell. Biol. 2:1362-1371(1982).
RN   [4]
RP   SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX   MEDLINE=83157109; PubMed=6219867;
RA   Monson J.M., McCarthy B.J.;
RT   "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT   evidence for insertions or deletions in gene coding sequences.";
RL   DNA 1:59-69(1981).
RN   [5]
RP   SEQUENCE OF 1442-1453 FROM N.A.
RX   MEDLINE=88124276; PubMed=3340560;
RA   Mooslehner K., Harbers K.;
RT   "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT   of the 3'-untranslated region.";
RL   Nucleic Acids Res. 16:773-773(1988).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08020; AAA88912.1; -.
DR   EMBL; X15896; CAA33904.1; -.
DR   EMBL; M14423; AAA37333.1; -.
DR   EMBL; M17491; AAA37334.1; -.
DR   EMBL; X06753; CAA29927.1; -.
DR   EMBL; K03036; AAA37332.1; -.
DR   EMBL; K03029; AAA37332.1; JOINED.
DR   EMBL; K03030; AAA37332.1; JOINED.
DR   EMBL; K03031; AAA37332.1; JOINED.
DR   EMBL; K03032; AAA37332.1; JOINED.
DR   EMBL; K03033; AAA37332.1; JOINED.
DR   EMBL; K03034; AAA37332.1; JOINED.
DR   EMBL; K03035; AAA37332.1; JOINED.
DR   PIR; A23982; A23982.
DR   MGD; MGI:88467; Colla1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
```

```
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     22
FT   PROPEP      23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      152   1207       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP    1208   1453       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN      29     87       VWFC.
FT   DOMAIN     152    167       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN     168   1181       TRIPLE-HELICAL REGION.
FT   DOMAIN    1182   1207       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD    56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1354   1354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE       734    736       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE      1082   1084       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT  1450   1450       A -> V (IN REF. 5).
SQ   SEQUENCE  1453 AA;  137944 MW;  3B802E535DF81808 CRC64;

[redacted]ment_scores:
            Quality:  450.00             Length:   1143
              Ratio:  0.947                Gaps:     62
 Percent Similarity:  41.557   Percent Identity: 26.072

alignment_block:
 US-09-673-254-1/rev x CA11_MOUSE   ..

 Align seg 1/1  to: CA11_MOUSE  from: 1  to: 1453

    2780 CCAGGT............CAATCACCTCGCAGGCTTCCTGCTGGCACGGC 2743
          ||||||           ::::::||||||   ||||||:::      ::
     167 ProGlyProMetGlyProSerGlyProArgGlyLeuProGlyProProGl 183

    2742 TGCTGCGGGA..............GCGGCTCGCGGGCGGGCGGCTGATC 2708
          :|||  |||              :::     |||:::  ::::::|
     183 yAlaProGlyProGlnGlyPheGlnGlyProProGlyGluProGlyGluP 200

    2707 CT............CACCTCGTCCGACGCGTACACCCAGGGCCG...... 2676
          ||            ::::      |||      |||||||||
     200 roGlyGlySerGlyProMetGlyProArgGlyProProGlyProProGly 216

    2675 .......................GATCGACCCGGACGACCTCAACGGCG 2650
                                ::::||||||||||||  :::||
     217 LysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProGlyGluAr 233

    2649 A......CCGTCACCGCTACAGCGCCGG......CCAGGCGTACGGCACG 2612
          |      |||:::|||         |||      ||||||:::      :
     233 gGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyL 250

     611 TCCAAACAGGCCAACATCATGACCGCGGC............GGAGGCCGC 2574
          ::      |||:::         ||||||               |||
     250 eu...ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAla 265

    2573 CAGGCGCTGGCCGGACGTGCTGGC........................... 2550
          ::::::   ||||||  ||||||
     266 LysGlyAspAlaGlyProAlaGlyProLysGlyGluProGlySerProGl 282

    2549 ................GGTCAGCTATCACCC................... 2535
                        ||||||::::::|||
     282 yGluAsnGlyAlaProGlyGlnMetGlyProArgGlyLeuProGlyGluA 299

    2534 .....CGGTGAGGTCCGCACCCGCATCGGACGGGGCACGGTCG....... 2497
              |||   ||||||  |||::::::   |||||||||:::
     299 rgGlyArgProGlyPro..ProGlyThrAlaGlyAlaArgGlyAsnAspG 315

    2496 ......................CCTCGTCCTACTTCCGGTTCAACCCCT 2470
                                |||   ||||||:::   :::   |||
     315 lyAlaValGlyAlaAlaGlyProProGlyProThrGlyProThrGlyPro 331

    2469 TCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTGGCGTCC 2420

          |||::::::                    |||   ::::::|||::::
     332 .ProGlyPhe.....................ProGlyAlaValGlyAlaL 341

    2419 GCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGACCGGCGGCT 2370
          :::::      |||   :::   :::|||      :::   |||   :::
     341 ysGlyGluAlaGlyProGlnGlyAlaArgGlySerGluGlyProGlnGly 357

    2369 GTCCCCGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCT 2320
          |||   ||||||        ||||||   ||||||:::
     358 ValArgGlyGlu..ProGlyProProGlyProAlaGly............ 369

    2319 GGGAGGCCGGCGCGGCCGCCGTCGGCGACACCGCGCACTGACGGC..... 2275
                   |||      |||::::::::::|||   :::||||||
     370 .........AlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnPr 383

    2274 .GGCGGCCCGCCCCGCCCGCATGTCCGTCTCATCCGCGAGATGTCCGTCT 2226
          ||| |||:::   |||:::   :::   :::   |||   :::|||:::
     383 oGly.AlaLysGlyAlaAsnGlyAlaProGlyIleAlaGlyAlaProGly 399

    2225 CATCCGCGAGCGCAGACGCTCGTGTGCCGATCCATCGAAAGGAACGATTC 2176
            |||   |||:::
     400 PheProGlyAlaArg................................... 404

    2175 GTGACCAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCC 2126
                        |||:::   |||:::|||              |
     405 .............GlyProSerGlyProGlnGly..............P 412

    2125 CGATGTCGCCGCCTCGGCCG.........ACTTCTACACCGGCCTGTTCG 2085
          ||   :::||||||   |||         ::::::   |||::::::
     412 roSerGlyProProGlyProLysGlyAsnSerGlyGluProGlyAlaPro 428

    2084 GCTGGACCGCCACCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACG 2035
          ::::::            :::|||         ||||||   |||
     429 GlyAsnLysGlyAspThrGlyAlaLysGlyGluProGlyAlaThrGlyVa 445

    2034 TTCAGCTCCGACGGGA............................... 2019
          ::::::|||   |||
     445 lGlnGlyProProGlyProAlaGlyGluGluGlyLysArgGlyAlaArgG 462

    2018 ..............AGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACA 1983
                       |||      |||      |||:::   |||:::
     462 lyGluProGlyProSerGlyLeuProGlyProProGlyGluArgGlyGly 478

    1982 CCCCCTACC.........................ACCGGCCGTACGGG 1960
          |||   :::                        :::|||      ||
     479 ProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysGl 495

    1959 CCCGGGAACGACCAGCACGGCATGCCGGCCATCTGGACCGTGTACTTCGC 1910
          ||||:::            |||      |||:::||||||      ||||
     495 yProSerGlyGluArgGlyAlaProGlyProAlaGlyProLysGlySerP 512

    1909 CACCGACGACGCCGACGCACTGACCAAGCGGGTCGAGACGGC........ 1868
          ||    :::      |||      :::||||||     |||
     512 ro.GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGl 528

    1867 .........................................GGGCGGC 1861
                                                    |||
     528 yLeuThrGlySerProGlySerProGlyProAspGlyLysThrGlyProP 545

    1860 GAGGTCATCATGACTCCGATGGACGTCCTCGGCCTCGG............ 1823
          |||        :::||||||||||||   |||
     545 roGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProProGly 561

    1822 ................................CCGGATGGCGGTCT 1809
                                         |||   |||
     562 AlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyThrAl 578

    1808 TCGCCGACCCCGCCGGGGCCGCGTTCGCGGTCTG......GCGCAAGGGA 1765
            :::|||   ::::::      |||||||||           |||:
     578 aGlyGluProGlyLysAlaGlyGluArgGlyLeuProGlyProProGlyA 595

    1764 GTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGCTC............ 1727
          ::   |||      |||   ||||||:::|||
```

```
 595 laValGlyProAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyAlaPro 611

1726 GGTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGG.........GGCCG 1686
     |||       |||   ||||||:::|||   ::::::       |||
 612 GlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyProAlaGlySe 628

1685 CCCGGGACTT......CTACCCGGCGACCCTCGGCCT......GGCTCCG 1648
      ||||||:::      |||||||||          |||         :::|
 628 rProGlyPheGlnGlyLeuProGlyProAlaGlyProProGlyGluAlaG 645

1647 GCCGACACCGGACTGAAGGGCGTCACCGACCCGGTCTGGCACATCGGTGA 1598
     ||:::    |||   :::||||||          :::         |||
 645 lyLys.ProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyPr 661

1597 CACACCGGTCGCCGGCACCCAG............GAGCTGGGCGTCACCG 1560
      :::   :::   |||   :::           |||   ||||||   |
 661 oSerGlyAlaArgGlyGluArgGlyPheProGlyGluArgGlyValGlnG 678

1559 GCGCGGTACGGCCGCACTGGGCCGTGCTGTTCGCCGTGCACGACTGCGAC 1510
     ||         ||| :::||||||
 678 lyProProGlyPro.AlaGlyPro.....................Ar 686

1509 GCGACGGTCCGGCGCGCCGTTGAACTCGGCGGCTCCGTCGAGAACGAGCC 1460
         ::::::|||||||||   |||:::::|||   :::   |||:::
 686 gGlyAsnAsnGlyAlaProGlyAsnAspGlyAlaLysGlyAspThrGlyA 703

1459 CGCCGACACGCCCAGGGGGCGGCGGGCGGACCTGCTCGACCCGCACGGGG 1410
       |||       ||||||
 703 laProGlyAlaProGly............................... 708

1409 CCGGCTTCTCGGTGGTCGAACTGCGGGAGGGGTACCCCGCGGCGGCGGGC 1360
                              ::::::|||:::|||      :::::
 709 ......................SerGlnGlyAlaProGlyLeuGlnGl 717

1359 GGTGCCTCGTGAGCGG..............CGAGGCGCCCCGGGTGGCC 1325
     :::|||   ||||||                 |||      |||   |
 717 yMetProGlyGluArgGlyAlaAlaGlyLeuProGlyProLysGlyAspA 734

1324 GTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGCCCGAGGT 1275
     ||      :::      :::     |||   :::          |||
 734 rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGly 750

1274 GCACGACGCCTT.....................CCGGGAGGCGGGCC 1249
     ||||||   |||                    |||:::|||:::||
 751 AlaArgGlyLeuThrGlyProIleGlyProProGlyProAlaGlyAlaPr 767

1248 CGGTCGTCGAGGTGAACGC...CCCCGCGGGCGGACCCGCCTGGGTCATC 1202
     ||||   :::||||||      |||   |||
 767 oGlyAspLysGlyGluAlaGlyProSerGly................... 777

1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
                |||   |||   ||||||      :::|||      :::
 778 ........ProProGlyProThrGlyAlaArgGlyAlaProGlyAspArg 791

1151 GGACCCCGA......CCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACG 1108
     |||           |||   |||      :::|||||||
 792 GlyGluAlaGlyProProGlyProAlaGlyPheAlaGlyProProGlyAl 808

1107 GTCTCGACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTG 1058
      :::   :::|||   :::   :::             :::      :::|
 808 aAspGlyGlnProGlyAlaLysGlyGluProGlyAspThrGlyValLysG 825

1057 GACGGCGAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAA 1008
     ||      |||||||||:::||||||          :::
 825 lyAspAlaGlyProProGlyProAla.........GlyProAlaGlyPro 838

1007 CCCGCGCCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCC 958
     |||:::|||   |||::::::   :::         :::      |||
 839 ProGlyProIleGlyAsnValGlyAlaProGlyProLysGlyProArgGl 855

 957 GGCTGCTCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGC. 909
      ||||||                        |||   |||   ||||||
 855 yAlaAla.....................GlyProProGlyAlaThrGlyP 865

 908 .....CGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCAT 864
           |||       |||:::      |||   |||:::   |||:::
 865 heProGlyAlaAlaGlyArgValGlyProProGlyProSer...GlyAsn 880

 863 CTGCGAGCTGCTCGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGG 814
                       ||||||      :::       ||||||   ||
 881 ..................AlaGlyProProGlyProProGlyProValGl 891

 813 CCGTCAGCGTTCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGG 764
     |::::::   :::::::|||   |||   :::
 891 yLysGluGlyGlyLysGlyProArgGlyGluThr................ 902

 763 GGTGACGGCACGGACCCTGCCGGGGGCGTGCCGGACACCTCGGC...CCT 717
               |||||||:::   ||||||   |||:::::::|||   |||
 903 ...........GlyProAla...GlyArgProGlyGluValGlyProPro 914

 716 GGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCCGGCGGAACGACACCC 667
     |||   |||:::|||   ::::::      ||||||:::          ||
 915 GlyProProGlyProAlaGlyGluLysGlySerProGlyAlaAspGlyPr 931

 666 CGACCATGACCCGCGTGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCG 617
     |      :::|||                                    |
 931 oAlaGlySerPro.....................................G 936

 616 GTCTCCGACGACCAGCTCGTCTACATGATCACCGGGCTCATCTTCGCCGG 567
     ||         |||      :::      :::|||:::   :::
 936 lyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGlyLeu 952

 566 CCA......CGACACCACCGGCTCCTTCCTGGGCTTCCTGCTCGCGGAGG 523
     |||      |||   :::|||   :::|||||||||||||:::   :::||
 953 ProGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSerGl 969

 522 T...CCTGGC......GGGCCGCCTCGCGGCGGATGCCGACGAGGACGCC 482
     |   ||||||      ||||||      ::::::   :::||||||
 969 yGluProGlyLysGlnGlyProSerGlySerSerGlyGluArgGlyProP 986

 481 GTCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTA 432
          |||:::         |||   |||   ||||||:::   :::
 986 roGlyProMetGlyProProGlyLeuAlaGlyProProGlyGluSerGly 1002

 431 CACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGG........... 393
     :::      :::      ::::::|||:::      |||
1003 ArgGluGlySerProGlyAlaGluGlySerProGlyArgAspGlyAlaPr 1019

 393 ..................................................... 393

1019 oGlyAlaLysGlyAspArgGlyGluThrGlyProAlaGlyProProGlyA 1036

 392 .....CGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACATCGA 348
               |||:::||||||      :::|||   :::|||   :::
1036 laProGlyAlaProGlyAlaProGlyProValGlyProAlaGlyLysAsn 1052

 347 GGGCACCAACACCGACGGCCGCCATCA.....................CG 319
     |||   :::   :::      |||:::
1053 GlyAspArgGlyGluThrGlyProAlaGlyProAlaGlyProIleGlyPr 1069

 318 ACGCCCCGCACGCCTTCCACCCGGACCGTC....CCTCGTGGCGGC.... 277
           ||||||   |||   ||||||      |||||||||
1069 oAlaGlyAlaArgGlyProAlaGlyProGlnGlyProArgGlyAspLysG 1086

 277 ..................................................... 277

1086 lyGluThrGlyGluGlnGlyAspArgGlyIleLysGlyHisArgGlyPhe 1102

 276 ....................GGCTCACCTTCGGCGACGGGCCGCACTAC 248
                         |||||||||::::::   |||:::
1103 SerGlyLeuGlnGlyProProGlySerProGlySerProGlyGluGlnGl 1119

 247 TGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCGGCGT 198
         ||||||:::|||:::|||:::      |||:::      :::
1119 yProSerGlyAlaSerGlyProAlaGlyProArgGlyProProGly.... 1134
```

```
     197 ACTGCGCAGCAGGTTCCCCGAGGCCCGAC............TGGCCGTGC 160
               :::||||||||||||      |||               |||
    1135 .....SerAlaGlySerProGlyLysAspGlyLeuAsnGlyLeuPro... 1148

     159 CGTACGACGAGTTGCGGTGGTGCCGGA...AGGGGGCCCAGACGGCGCGG 113
                       |||      |||      |||||||||   |||
    1149 ..............GlyProIleGlyProProGlyProArgGlyArgTh 1160

     112 CTCACCGAACTGCCCGTCTGGCTGCGCTGATCCGCCGGCCGCGACCGGC. 64
           :::   :::   |||:::|||         ||||||:::      |||
    1160 rGlyAspSerGlyProAlaGlyProProGlyProProGlyProProGlyP 1177

      63 ..ACGGGACCGCCCACC 49
               |||||||||:::
    1177 roProGlyProProSer 1182

seq_name: SwissProt_40:CA13_MOUSE

seq_documentation_block:
ID   CA13_MOUSE     STANDARD;      PRT;  1464 AA.
AC   P08121; Q61429; Q9CRN7;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(III) chain precursor.
GN   COL3A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL X DBA; TISSUE=Embryo;
RX   MEDLINE=95011609; PubMed=7926795;
RA   Toman D., de Crombrugghe B.;
RT   "The mouse type-III procollagen-encoding gene: genomic cloning and
RT   complete DNA sequence.";
RL   Gene 147:161-168(1994).
RN   [2]
RP   SEQUENCE OF 1-488 FROM N.A.
RX   MEDLINE=88167858; PubMed=3443309;
RA   Wood L., Theriault N., Vogeli G.;
RT   "Complete nucleotide sequence of the N-terminal domains of the murine
RT   alpha-1 type-III collagen chain.";
RL   Gene 61:225-230(1987).
RN   [3]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=85131189; PubMed=3972847;
RA   Liau G., Mudryj M., de Crombrugghe B.;
RT   "Identification of the promoter and first exon of the mouse alpha 1
RT   (III) collagen gene.";
     J. Biol. Chem. 260:3773-3777(1985).
     [4]
     SEQUENCE OF 810-1464 FROM N.A.
     STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [5]
RP   SEQUENCE OF 1442-1464 FROM N.A.
RC   STRAIN=C57BL;
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaeranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC       LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC       ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC       O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52046; CAA36279.1; -.
DR   EMBL; M18933; AAA37338.1; -.
DR   EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR   EMBL; AK019448; BAB31724.1; -.
DR   EMBL; X57983; CAA41048.1; -.
DR   PIR; A22287; A22287.
DR   PIR; A27353; A27353.
DR   PIR; S16373; S16373.
DR   MGD; MGI:88453; Col3a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     23       BY SIMILARITY.
FT   PROPEP      24    154       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      155   1203       COLLAGEN ALPHA 1(III) CHAIN.
FT   PROPEP    1204   1464       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN      31     90       VWFC.
FT   DOMAIN     155    169       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN     170   1195       TRIPLE-HELICAL REGION.
FT   DOMAIN    1196   1464       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD   262    262       O-LINKED (GAL. . .) (BY SIMILARITY).
FT   MOD_RES    262    262       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    283    283       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    859    859       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    976    976       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES   1093   1093       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES   1105   1105       HYDROXYLATION (BY SIMILARITY).
FT   DISULFID  1195   1195       INTERCHAIN (BY SIMILARITY).
FT   DISULFID  1196   1196       INTERCHAIN (BY SIMILARITY).
SQ   SEQUENCE   1464 AA;  138944 MW;  2104EC27A886090B CRC64;


alignment_scores:
           Quality:  450.00            Length:   1187
             Ratio:  0.911               Gaps:     77
 Percent Similarity:  41.618   Percent Identity: 27.043
```

```
alignment_block:
 US-09-673-254-1/rev x CA13_MOUSE   ..

 Align seg 1/1  to: CA13_MOUSE  from: 1  to: 1464

   2844 CCAGCAACGCCGGCGGCATGTTCTGGTCGCGCACCACGACCCAGGACG.. 2797
        |||      |||   :::   ||||||:::         |||   :::
    180 ProProGlyProProGlySerSerGlyHisProGlySerProGlySerPr 196

   2796 .GGTTCGAGGCCACCATCCAGGTCAATCACCTCGCAGGCTTCCTGCTGGC 2748
         |||   :::   ||| ||||||:::      :::      |||||||||
    196 oGlyTyrGlnGlyPro.ProGlyGluProGlyGlnAlaGlyProAlaGly 212

   2747 ACGGCTGCTGCGGGAGCGGCTCGCGGGCGGGCGGCTGATCCTCACCTCGT 2698
           :::      :::|||   |||   |||
    213 ProProGlyProProGlyAlaLeuGlyProAla..................  223

   2697 CCGACGCGTACACCCAGGGCCG.....................GATCGAC 2669
                       ||||||                      ||||
    224 ................GlyProAlaGlyLysAspGlyGluSerGlyArgP 235

   2668 CCGGACGACCTCAACGGCGA......CCGTCACCGCTACAGCGCCGGCCA 2625
        ||||||||||||   :::|||      |||:::|||             |||
    235 roGlyArgProGlyGluArgGlyLeuProGlyPro............Pro 247

   2624 GGCGTACGGCACGTCCAAACAGGCCAACATCATGACCGCGGCGGAGGCCG 2575
        |||::::::
    248 GlyIleLys....................................... 250

   2574 CCAGGCGCTGGCCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTC 2525
                    |||   ||||||          |||      |||
    251 .............GlyProAlaGlyMetProGlyPhePro......GlyM 261

   2524 CGCACCCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCAA 2475
              ||||||         ||||||
    261 etLysGlyHisArgGlyPheAspGlyArgAsnGlyGluLysGlyGluThr 277

   2474 CCCCTTCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTGG 2425
           |||:::|||:::||||||:::                        ||
    278 GlyAlaProGlyLeuLysGlyGluAsn.....................Gl 287

   2424 CGTCCGCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGACCGG 2375
        |:::   :::   :::|||:::                  :::   |||
    287 yLeuProGlyAspAsnGlyAlaProGlyPro.........MetGlyProA 301

   2374 CGGCTGTCCCCGGTGAGCGGCC.CGACCGCCGACGCCG............ 2338
          ::::::||||||||||||   ||||||      |||
    301 rgGlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGly 317

   2337 ..............GCCTCGCGGCGAAGCTCTGGGAGGCCGGCGCGGCC 2303
                      ::::::|||   |||:::|||:::|||:::
    318 AlaArgGlyAsnAspGlyAlaArgGlySerAspGlyGlnProGly..... 332

   2302 GCCGTCGGCGACACCGCGCACTGACGGC......GGCGGCCCGCCCCGCC 2259
         |||   :::   |||   |||   |||      |||:::|||    |||
    333 .ProProGlyProProGlyThrAlaGlyPheProGlySerPro.GlyAla 348

   2258 CGCATGTCCGTCTCATCCGCGAGATGTCCGTCTCATCCGCGAGCGCAGAC 2209
        :::   :::|||:::   |||      |||:::
    349 LysGlyGluValGlyProAlaGlySerProGly.................. 359

   2208 GCTCGTGTGCCGATCCATCGAAAGGAACGATTCGTGACCAGGTTCGCGCC 2159
                            :::::::::   ||||||:::|||
    360 .........................SerAsnGlySerProGlyGlnArgG 368

   2158 CGGCGCCC...CCGCATGGTTCGACCTCGGTTCGC...CCGATGTCGCCG 2115
           |||   |||:::|||   ::::::::::::   |||   :::|||
    368 lyGluProGlyProGlnGlyHisAlaGlyAlaGlnGlyProProGlyPro 384

   2114 CCTCGGCCGACTTCTACACCGGCCTGTTCGGCTGGACCGCCACCGTGGTC 2065
        |||      ::::::::::|||:::   ::::::          |||   ::
    385 ProGlyAsnAsnGlySerProGlyGlyLysGlyGluMetGlyProAlaGl 401

   2064 AGCGACCCGGGCGCCGGGGGATACACGACGTTCAGCTCCGACGGGAAGCT 2015
        :         |||||||||                           |||:::
    401 yIleProGlyAlaProGly.................LeuIleGlyAlaA 412

   2014 GGTCGCCGCGGTCGCCCGCC...ACCAGATCGACACCCCCTACCACCGGC 1968
          :::|||   :::||||||   |||::::::   |||   |||   |||
    412 rgGlyProProGlyProAlaGlyThrAsnGlyIleProGlyThrArgGly 428

   1967 CGTACGGGC...CCGGGAACGACCAGCACGGCATGCCGGCCATCT..... 1926
           ::|||   ||||||   :::::::::      :::|||:::
    429 ProSerGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyAlaAr 445

   1925 ..........GGACCGTGTACTTCGCCA..................... 1908
                   |||   :::   ||||||
    445 gGlyGluArgGlyGluAlaGlySerProGlyIleProGlyProLysGlyG 462

   1907 ....................CCGACGACGCCG.................. 1896
                            |||      |||
    462 luAspGlyLysAspGlySerProGlyGluProGlyAlaAsnGlyLeuPro 478

   1895 ............ACGCACTGACCAAGCGGGTCGAGACGGC.......... 1868
                       :::   |||||||||   ||| |||
    479 GlyAlaAlaGlyGluArgGlyProSerGlyPheArg.GlyProAlaGlyP 495

   1867 ..............GGGCGGCGAGGTCATCATGACTCCGATGGACGTCCT 1832
                     |||::::::|||         :::   |||   |||
    495 roAsnGlyIleProGlyGluLysGlyProProGlyGluArgGlyGlyPro 511

   1831 CGGCCTCGG...CCGGATGGCGGTCTTCGCCGACCCCGCCGGGGC..... 1790
           |||      |||   |||      :::|||   |||
    512 GlyProAlaGlyProArgGlyValAlaGlyGluProGlyArgAspGlyTh 528

   1789 ............CGCGTTCGCGGTCTGGCGCAAGGGAGTCATGGAGGGC 1753
                       :::||||||::      ||||||   ||||||
    528 rProGlyGlyProGlyIleArgGlyMet...ProGlySerProGlyGlyP 544

   1752 GCGGAGGTGACGGGCGT.....................GCCCGGCTCGGT 1724
          |||:::||||||:::                     ::::::   |||
    544 roGlyAsnAspGlyLysProGlyProProGlySerGlnGlyGluSerGly 560

   1723 CGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCGGGACTTCT 1674
        |||   |||      |||            ||||||:::||||||:::::
    561 ArgProGlyProProGlyProSerGlyProArgGlyGlnProGlyValMe 577

   1673 A.............................................. 1673
        :
    577 tGlyPheProGlyProLysGlyAsnAspGlyAlaProGlyLysAsnGlyG 594

   1672 ...........CCCGGCGACCCTCGGCCTGGCTCCGGCCGACACCGGACT 1634
                  ||||||   ||| ||||||   |||:::   :::|||
    594 luArgGlyGlyProGlyGlyPro.GlyLeu...ProGlyProAlaGlyLy 609

   1633 GAAGGGCGTCACCGACCCGGTCTGGCACATCGGTGACACACCGGTCGCCG 1584
         :::|||   |||   |||            |||   |||      ||||
    609 sAsnGlyGluThrGlyProGlnGlyProProGlyProThrGlyProAlaG 626

   1583 GCACCCAGGAGCTGGGCGTCACCGGCGCGGTACGGCCGCACTGGGCCGTG 1534
        ||      :::   |||   :::|||         |||::: |||
    626 ly......AspLysGlyAspSerGlyProProGlyProGln.GlyLeuGl 640

   1533 CTGTTCGCCGTGCACGACTGCGACGCGACGGTCCGGCGCGCCGTTGAACT 1484
           |||   |||               :::|||   |||   ::::
    640 nGlyIleProGlyThr..............GlyGlyProProGlyGluA 652

   1483 CGGCGGCTCCGTCGAGAACGAGCCCG.................CCGACA 1452
        :::::   |||::::::   :::|||                |||
    652 snGlyLysProGlyGluProGlyProLysGlyGluValGlyAlaProGly 668

   1451 CGCCCAGGGGGCGGCGGGCGGACCTGCTCGACCCGCACGGGGCCGGCTTC 1402
           |||||||||   |||   :::                :::|||:::::
    669 AlaProGlyGlyLysGlyAspSer............GlyAlaProGlyGl 681
```


1401 TCGGTGGTCGAACTGCGGGAGGGGTACCCCGCGGCGGCGGGCGGTGCCTC 1352
681 u.....ArgGlyProProGlyThrAlaGlyIleProG 692
1351 GTGAGCGGCGAGGCGCC.....CCGGGTGGCCGTCGACCC 1317
692 lyAlaArgGlyGlyAlaGlyProProGlyProGluGlyGlyLysGlyPro 708
1316 GTTCGCGTGTCCCATGATGACCATGCAGCGCAAGCCCGAGGTGCACGACG 1267
709 AlaGlyProProGlyProProGlyAlaSerGlySerProGlyLeuGlnGl 725
1266 CCTTCCGGGA.....GGCGGGCCCGGTCGTCGAGGTGAACGCCCCCGCG 1223
725 yMetProGlyGluArgGlyGlyProGlySerProGly.....ProLysG 740
1222 GGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCGCGAGGTGCT 1173
740 lyGlu.....LysGlyGluProGlyGlyAla 748
1172 GGCCGATCC.....CCGGTTCGTGAAGGACCCCGA..... 1143
749 GlyAlaAspGlyValProGlyLysAspGlyProArgGlyProAlaGlyPr 765
1142 .....CCTCGCCCCCGCCGCCTGGCGGGG.....GGTG 1115
765 oIleGlyProProGlyProAlaGlyGlnProGlyAspLysGlyGluGlyG 782
1114 GACGACGGTCT.....CGACATCCCCGTTCCGGAGCTGCGTCC 1077
782 lySerProGlyLeuProGlyIleAlaGlyProArgGlyGlyProGly... 797
1076 GTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCGCCGCA 1027
798 .....GluArgGlyGluHisGlyProProGlyProAlaGlyPh 810
1026 TCCACGCACCTGCGTTCAA..... 1008
810 eProGlyAlaProGlyGlnAsnGlyGluProGlyAlaLysGlyGluArgG 827
1007 .....CCCGCGCCGGCTGGCCGA 990
827 lyAlaProGlyGluLysGlyGluGlyGlyProProGlyProAlaGlyPro 843
989 GCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCG 940
844 ThrGlySerSer.....GlyProAlaGly..... 851
939 CCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTC 890
852 .....ProProGlyProGlnGlyValLysGlyGluArg.... 862
889 GCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGT 840
863 .....GlySerProGlyGlyProGlyThrAlaGlyPheProGly 875
839 CACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCG 790
876 GlyArgGlyLeuProGlyProProGlyAsnAsnGlyAsnProGly..... 890
789 GCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGG 740
891 .ProProGlyProSerGlyAlaProGly...LysAspGlyPro...ProG 905
739 GGCGTGCCGGACACCTCGGC...CCTGGAGAGCCT.....GCTCCT 702
905 lyProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaGlyPro 921
701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGA..... 664
922 LysGlyAspAlaGlyGlnProGlyGluLysGlyProProGlyAlaGlnGl 938
663 .....CCATGACCC.....GCGTGCTGTACGAGCGCGC 636
938 yProProGlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaA 955
635 GCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCA 586
955 rg.....GlyLeuAlaGlyProProGlyMetProGlyProArgGlySer 969
585 CCGGGC.....TCATCTTCGCCGGCCACGACA.. 559
970 ProGlyProGlnGlyIleLysGlyGluSerGlyLysProGlyAlaSerGl 986
558 .....CCACCGGCTCCTTCCTGGGCTTCCTGCTCGC 528
986 yHisAsnGlyGluArgGlyProProGlyProGln.GlyLeuProGlyGln 1002
527 GGAGGTCCTGGCGGGCCGCCT...CGCGGCGGATGCCGACGAGGACGCCG 481
1003 ProGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGl 1019
480 TCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTAC 431
1019 yGlnProGlyArgAspGlySer.....ProGlyGlyLys.... 1030
430 ACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGCT 381
1031 .....GlyAsp...ArgGlyGluAsnGly 1037
380 GCCCCGCGGAGCGCCGGTGCTGGTGGACAT..... 351
1038 SerProGlyAlaProGlyAlaProGlyHisProGlyProProGlyProVa 1054
350 .....CGAGGGCACCAACACCGACGGCCGC 326
1054 lGlyProSerGlyLysSerGlyAspArgGlyGluThrGlyProAlaGlyP 1071
325 CATCACGACGCCCC.....GCACGCCTTCCACCCGGACCGTC. 289
1071 roSerGlyAlaProGlyProAlaGlyAlaArgGlyAlaProGlyProGln 1087
288 ...CCTCGTGGCGGC..... 277
1088 GlyProArgGlyAspLysGlyGluThrGlyGluArgGlySerAsnGlyIl 1104
276 .....GGCTCACCTT 267
1104 eLysGlyHisArgGlyPheProGlyAsnProGlyProProGlySerProG 1121
266 CGGCGACGGGCCGCACTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGT 217
1121 lyAlaAlaGlyHisGlnGlyAlaIleGlySerProGlyProAlaGlyPro 1137
216 CGCGCACGATGATCGGCGTAC.....TGCGCAGCAGG 185
1138 ArgGlyProValGlyProHisGlyProProGlyLysAspGlyThrSerGl 1154
184 TTCCCCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCG 135
1154 yHisProGlyProIleGlyPro..... 1161
134 GAAGGGGGCCCAGA.....CGGCGCGGCTCACCGAACTGCCCG 97
1162 ..ProGlyProArgGlyAsnArgGlyGluArgGlySerGluGlySerPro 1177
96 TCT.....GGCTGCGCTGATCCGCCGGCCGCGACCGGCACGGGACCGCC 53
1178 GlyHisProGlyGlnProGlyProPro.....GlyProPr 1189
52 CACCGCGCACCGCGCGGTGGGCGGTCCCGTGCCGTGTACCCGGTGCGGA 4
1189 o.....GlyAlaProGlyProCys.....CysGly 1197

```
OM of: US-09-673-254-1 to: SPTREMBL_19:*   out_format : pfs

Date: Jun 11, 2002  8:19 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09673254/runat_11062002_114213_1012/app_query.fasta_1.2973
 -DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=n2p.rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -[illegible]LOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -[illegible]NS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -[illegible]_MAX=100 -THR_MIN=0 -ALIGN=7 -MODE=LOCAL -OUTFMT=pfs
 -[illegible]M=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09673254_@CGN1_1_360 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-673-254-1
Query length: 2870
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 134.250000


score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
sp_bacteria:Q9ZAU3    - 2171.00 1722.23  3.3e-88   415  ! Q9zau3 streptomyces peucetius.
sp_bacteria:Q59971    - 2087.00 1655.77  1.6e-84   422  ! Q59971 streptomyces sp. cytochr
sp_bacteria:Q93MI2    - 2062.00 1636.07  2.0e-83   420  ! Q93mi2 streptomyces peucetius s
sp_bacteria:Q9ZAU2    - 1465.00 1168.27  3.6e-57   275  ! Q9zau2 streptomyces peucetius.
sp_bacteria:Q55078    - 1385.00 1105.11  1.2e-53   275  ! Q55078 streptomyces sp. orfa pr
sp_bacteria:Q9ZAU1    - 1014.00  811.84  2.4e-37   287  ! Q9zau1 streptomyces peucetius.
sp_bacteria:Q53877    -  980.00  785.00  7.6e-36   287  ! Q53877 streptomyces sp. c5. ket
sp_bacteria:Q55077    -  975.00  784.27  1.2e-35   195  ! Q55077 streptomyces sp. daunomy
sp_virus:Q99307       -  538.00  426.28  2.3e-16   924  ! Q99307 epstein-barr virus (stra
sp_vertebrate:Q9YIB4  -  506.00  397.25  6.0e-15  1450  ! Q9yib4 cynops pyrrhogaster (ja
sp_virus:Q9Q5K9       +  486.50  389.11  4.0e-14   608  ! Q9q5k9 herpesvirus papio. ntr.
sp_human:O60382       -  486.50  380.09  4.4e-14  1791  ! O60382 homo sapiens (human). k
sp_human:Q9UQ35       -  484.50  374.93  5.5e-14  2752  ! Q9uq35 homo sapiens (human). r
sp_human:O15038       -  484.00  378.16  5.6e-14  1783  ! O15038 homo sapiens (human). k
sp_human:O15038       +  479.00  374.21  9.3e-14  1783  ! O15038 homo sapiens (human). k
sp_human:O60382       +  479.00  374.17  9.3e-14  1791  ! O60382 homo sapiens (human). k
sp_human:Q9UQ35       +  479.00  370.59  9.6e-14  2752  ! Q9uq35 homo sapiens (human). r
sp_human:Q9UHA8       -  478.00  371.31  1.0e-13  2296  ! Q9uha8 homo sapiens (human). s
sp_vertebrate:O93251  -  476.50  373.99  1.2e-13  1445  ! O93251 rana catesbeiana (bull
sp_invertebrate:Q9VPG1 -  475.50  369.38  1.4e-13  2284  ! Q9vpg1 drosophila melanogast
sp_human:Q9UQ36       -  473.00  372.27  1.7e-13  1275  ! Q9uq36 homo sapiens (human). r
sp_bacteria:Q9RK97    -  472.50  385.05  1.6e-13   263  ! Q9rk97 streptomyces coelicolor.
sp_virus:O41971       -  470.00  374.07  2.2e-13   774  ! O41971 murid herpesvirus 4. hyp
sp_human:Q9UQ39       +  468.00  370.73  2.7e-13   956  ! Q9uq39 homo sapiens (human). rn
sp_bacteria:Q9F2Q0    -  467.00  376.55  2.8e-13   433  ! Q9f2q0 streptomyces coelicolor.
sp_virus:O41971       +  460.50  366.57  5.7e-13   774  ! O41971 murid herpesvirus 4. hyp
sp_human:Q9UQ40       -  459.00  361.30  6.9e-13  1262  ! Q9uq40 homo sapiens (human). r
sp_vertebrate:Q9IB91  -  458.50  359.77  7.3e-13  1447  ! Q9ib91 xenopus laevis (africa
sp_human:O76045       -  458.00  359.29  7.7e-13  1461  ! O76045 homo sapiens (human). p
sp_rodent:Q63079      -  454.50  356.58  1.1e-12  1453  ! Q63079 rattus norvegicus (rat)
sp_invertebrate:P91365 -  454.00  352.60  1.2e-12  2232  ! P91365 caenorhabditis elega
sp_human:Q9UHA8       +  452.50  351.18  1.4e-12  2296  ! Q9uha8 homo sapiens (human). s
sp_human:Q9UQ39       -  452.00  358.09  1.4e-12   956  ! Q9uq39 homo sapiens (human). rn
sp_virus:O41972       -  451.50  360.23  1.4e-12   706  ! O41972 murid herpesvirus 4. hyp
sp_virus:O41973       -  451.00  359.59  1.5e-12   727  ! O41973 murid herpesvirus 4. hyp
sp_human:Q9UQ40       +  450.50  354.59  1.6e-12  1262  ! Q9uq40 homo sapiens (human). r
sp_human:Q9UQ36       +  449.00  353.32  1.9e-12  1275  ! Q9uq36 homo sapiens (human). r
sp_virus:Q99307       +  448.50  355.62  2.0e-12   924  ! Q99307 epstein-barr virus (stra
sp_bacteria:O52569    -  444.50  357.40  2.8e-12   511  ! O52569 amycolatopsis mediterra
sp_virus:Q9Q5K9       -  443.50  355.16  3.1e-12   608  ! Q9q5k9 herpesvirus papio. ntr.
sp_virus:O41972       +  437.50  349.18  5.8e-12   706  ! O41972 murid herpesvirus 4.
sp_vertebrate:Q90YJ0  -  436.00  342.57  7.1e-12  1352  ! Q90yj0 brachydanio rerio (z
sp_virus:Q69340       +  434.50  338.30  8.5e-12  1958  ! Q69340 pseudorabies virus.
sp_human:Q9UF83       -  430.50  345.29  1.2e-11   580  ! Q9uf83 homo sapiens (human).
sp_invertebrate:Q9VE45 -  430.00  340.78  1.3e-11   950  ! Q9ve45 drosophila melanoga

seq_name: sp_bacteria:Q9ZAU3

seq_documentation_block:
ID   Q9ZAU3      PRELIMINARY;      PRT;   415 AA.
AC   Q9ZAU3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CYTOCHROME P450 129A2 (DAUNORUBICIN C-14 HYDROXYLASE).
GN   CYP129A2 OR DOXA.
OS   Streptomyces peucetius.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1950;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29050;
RX   MEDLINE=99084971; PubMed=9864344;
RA   Lomovskaya N., Otten S.L., Doi-Katayama Y., Fonstein L., Liu X.-C.,
RA   Takatsu T., Inventi A., Filippi S., Torti F., Colombo A.L.,
RA   Hutchinson C.R.;
RT   "Doxorubicin overproduction in Streptomyces peucetius: cloning and
RT   characterization of the dnrU ketoreductase and dnrV genes and the doxA
RT   cytochrome P-450 hydroxylase gene.";
RL   J. Bacteriol. 181:305-318(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U77891; AAD04715.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 2.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   415 AA;  45095 MW;  CCF873824BEB6CCF CRC64;


alignment_scores:
            Quality: 2171.00              Length:    415
              Ratio:  5.231                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.759

alignment_block:
 US-09-673-254-1/rev x Q9ZAU3   ..

 Align seg 1/1  to: Q9ZAU3  from: 1  to: 415

    1330 GTGGCCGTCGACCCGTTCGCGTGTCCCATGATGACCATGCAGCGCAAGCC 1281
         :::|||||||||||||||||||||||||||||||||||||||||||||||
       1 MetAlaValAspProPheAlaCysProMetMetThrMetGlnArgLysPr 17

    1280 CGAGGTGCACGACGCCTTCCGGGAGGCGGGCCCGGTCGTCGAGGTGAACG 1231
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 oGluValHisAspAlaPheArgGluAlaGlyProValValGluValAsnA 34

    1230 CCCCCGCGGGCGGACCCGCCTGGGTCATCACCGATGACGCCCTCGCCCGC 1181
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 laProAlaGlyGlyProAlaTrpValIleThrAspAspAlaLeuAlaArg 50

    1180 GAGGTGCTGGCCGATCCCCGGTTCGTGAAGGACCCCGACCTCGCCCCCGC 1131
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GluValLeuAlaAspProArgPheValLysAspProAspLeuAlaProAl 67

    1130 CGCCTGGCGGGGGGTGGACGACGGTCTCGACATCCCCGTTCCGGAGCTGC 1081
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 aAlaTrpArgGlyValAspAspGlyLeuAspIleProValProGluLeuA 84

    1080 GTCCGTTCACGCTCATCGCCGTGGACGGCGAGGCCCACCGGCGCCTGCGC 1031
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 rgProPheThrLeuIleAlaValAspGlyGluAlaHisArgArgLeuArg 100

1030 CGCATCCACGCACCTGCGTTCAACCCGCGCCGGCTGGCCGAGCGGACGGA 981
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 ArgIleHisAlaProAlaPheAsnProArgArgLeuAlaGluArgThrAs 117

 980 TCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGCCT 931
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 pArgIleAlaAlaIleAlaGlyArgLeuLeuThrGluLeuAlaAspAlaS 134

 930 CCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCAC 881
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 erGlyArgSerGlyLysProAlaGluLeuIleGlyGlyPheAlaTyrHis 150

 880 TTCCCGCTGTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCC 831
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 PheProLeuLeuValIleCysGluLeuLeuGlyValProValThrAspPr 167

 830 GGCGATGGCCCGCGAGGCCGTCAGCGTTCTCAAGGCACTCGGCCTCGGCG 781
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 oAlaMetAlaArgGluAlaValSerValLeuLysAlaLeuGlyLeuGlyG 184

 780 GCCCGCAGAGCGGCGGGGGTGACGGCACGGACCCTGCCGGGGGCGTGCCG 731
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 lyProGlnSerGlyGlyGlyAspGlyThrAspProAlaGlyGlyValPro 200

 730 GACACCTCGGCCCTGGAGAGCCTGCTCCTCGAAGCCGTGCACTCAGCCCG 681
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 AspThrSerAlaLeuGluSerLeuLeuLeuGluAlaValHisSerAlaAr 217

 680 GCGGAACGACACCCCGACCATGACCCGCGTGCTGTACGAGCGCGCGCAGG 631
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 gArgAsnAspThrProThrMetThrArgValLeuTyrGluArgAlaGlnA 234

 630 CCGAGTTCGGCTCGGTCTCCGACGACCAGCTCGTCTACATGATCACCGGG 581
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 laGluPheGlySerValSerAspAspGlnLeuValTyrMetIleThrGly 250

 580 CTCATCTTCGCCGGCCACGACACCACCGGCTCCTTCCTGGGCTTCCTGCT 531
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuIlePheAlaGlyHisAspThrThrGlySerPheLeuGlyPheLeuLe 267

 530 CGCGGAGGTCCTGGCGGGCCGCCTCGCGGCGGATGCCGACGAGGACGCCG 481
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 uAlaGluValLeuAlaGlyArgLeuAlaAlaAspAlaAspGluAspAlaV 284

 480 TCTCCCGGTTCGTGGAGGAGGCGCTGCGCTACCACCCGCCGGTGCCCTAC 431
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 alSerArgPheValGluGluAlaLeuArgTyrHisProProValProTyr 300

 430 ACGTTGTGGAGGTTCGCTGCCACGGAGGTGACCATCGGCGGCGTCCGGCT 381
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 ThrLeuTrpArgPheAlaAlaThrGluValThrIleGlyGlyValArgLe 317

 380 GCCCCGCGGAGCGCCGGTGCTGGTGGACATCGAGGGCACCAACACCGACG 331
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 uProArgGlyAlaProValLeuValAspIleGluGlyThrAsnThrAspG 334

 330 GCCGCCATCACGACGCCCCGCACGCCTTCCACCCGGACCGTCCCTCGTGG 281
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 lyArgHisHisAspAlaProHisAlaPheHisProAspArgProSerTrp 350

 280 CGGCGGCTCACCTTCGGCGACGGGCCGCACTACTGCATCGGGGAGCAGCT 231
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 ArgArgLeuThrPheGlyAspGlyProHisTyrCysIleGlyGluGlnLe 367

 230 CGCCCAGCTGGAGTCGCGCACGATGATCGGCGTACTGCGCAGCAGGTTCC 181
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 uAlaGlnLeuGluSerArgThrMetIleGlyValLeuArgSerArgPheP 384

 180 CCGAGGCCCGACTGGCCGTGCCGTACGACGAGTTGCGGTGGTGCCGGAAG 131
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 384 roGluAlaArgLeuAlaValProTyrAspGluLeuArgTrpCysArgLys 400

 130 GGGGCCCAGACGGCGCGGCTCACCGAACTGCCCGTCTGGCTGCGC 86
     ||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAlaGlnThrAlaArgLeuThrGluLeuProValTrpLeuArg 415

seq_name: sp_bacteria:Q59971

seq_documentation_block:
ID   Q59971      PRELIMINARY;      PRT;   422 AA.
AC   Q59971;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CYTOCHROME P450 129A1 (DAUNOMYCIN C-14 HYDROXYLASE).
GN   CYP129A1 OR DOXA.
OS   Streptomyces sp.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1931;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C5;
RX   MEDLINE=96236066; PubMed=8655529;
RA   Dickens M.L., Strohl W.R.;
RT   "Cloning, sequencing, and analysis of aklaviketone reductase from
RT   Streptomyces sp. strain C5.";
RL   J. Bacteriol. 178:3384-3388(1996).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U50973; AAB08049.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 2.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW   Electron transport; Heme; Membrane; Monooxygenase; Oxidoreductase.
FT   BINDING     369    369       HEME (BY SIMILARITY).
SQ   SEQUENCE   422 AA;  46096 MW;  F4EADECA1D159052 CRC64;


alignment_scores:
            Quality: 2087.00              Length:    422
              Ratio:  5.041                Gaps:      0
 Percent Similarity:  98.104   Percent Identity: 94.550

alignment_block:
 US-09-673-254-1/rev x Q59971   ..

 Align seg 1/1  to: Q59971  from: 1  to: 422

    1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
         :::|||||||||||||||||||||||||||||||||||:::||||||||
       1 MetSerGlyGluAlaProArgValAlaValAspProPheSerCysProMe 17

    1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

    1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpValIle 50

    1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ThrAspAspAlaLeuAlaArgGluValLeuAlaAspProArgPheValLy 67

    1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCG 1102
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
      67 sAspProAspLeuAlaProThrAlaTrpArgGlyValAspAspGlyLeuA 84

    1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100
```

```
1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
     |||   ||||||||||||||||||||||||||||||||||||||||||||
 101 GluAspHisArgArgLeuArgArgIleHisAlaProAlaPheAsnProAr 117

1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
     |||||||||||||||||||||||||||||||||||   |||||||
 117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaAspArgLeuL 134

 951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
     ||||||||||||||||:::|||   |||||||||:::|||||||||||||
 134 euThrGluLeuAlaAspSerSerAspArgSerGlyGluProAlaGluLeu 150

 901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 IleGlyGlyPheAlaTyrHisPheProLeuLeuValIleCysGluLeuLe 167

 851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
     ||||||||||||||||||||||||||||||||||||||||||:::||||
 167 uGlyValProValThrAspProAlaMetAlaArgGluAlaValGlyValL 184

 801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
     ||||||||||||||||||||||||||||||||:::|||||||||||||||
 184 euLysAlaLeuGlyLeuGlyGlyProGlnSerAlaGlyGlyAspGlyThr 200

 751 GACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
     ||||||||||||   ||||||||||||||||||||||||||||||||||
 201 AspProAlaGlyAspValProAspThrSerAlaLeuGluSerLeuLeuLe 217

 701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
     ||||||||||||:::|||||||||:::||||||   ||||||||||||||
 217 uGluAlaValHisAlaAlaArgArgLysAspThrArgThrMetThrArgV 234

 651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 alLeuTyrGluArgAlaGlnAlaGluPheGlySerValSerAspAspGln 250

 601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 LeuValTyrMetIleThrGlyLeuIlePheAlaGlyHisAspThrThrGl 267

 551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 284

 501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
     ||||||||||||   ||||||:::|||||||||||||||||||||||||||
 284 laAspAlaAspGlyAspAlaIleSerArgPheValGluGluAlaLeuArg 300

 451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
     :::|||||||||||||||||||||||||||||||||||||||||||||||
 301 HisHisProProValProTyrThrLeuTrpArgPheAlaAlaThrGluVa 317

 401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
     |:::|||   |||||||||||||||||||||||||||||||||||||||||
 317 lValIleArgGlyValArgLeuProArgGlyAlaProValLeuValAspI 334

 351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 350

 301 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
     ||||||||||||||||   ||||||||||||||||||||||||||||||||
 351 HisProAspArgProSerArgArgArgLeuThrPheGlyAspGlyProHi 367

 251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 384

 201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
     ||||||||||||||||||||||:::|||||||||||||||||||||||:::
 384 lyValLeuArgSerArgPheProGlnAlaArgLeuAlaValProTyrGlu 400

 151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||:::||
 401 GluLeuArgTrpCysArgLysGlyAlaGlnThrAlaArgLeuThrAspLe 417

 101 GCCCGTCTGGCTGCGC 86
     ||||||||||||||||
 417 uProValTrpLeuArg 422
```

```
seq_name: sp_bacteria:Q93MI2

seq_documentation_block:
ID   Q93MI2      PRELIMINARY;      PRT;   420 AA.
AC   Q93MI2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DAUNOMYCIN C-14 HYDROXYLASE.
GN   DOXA.
OS   Streptomyces peucetius subsp. caesius.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=55158;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27952;
RA   Hong Y.-S., Kim H.S., Lee J.-H., Kim K.-W., Lee J.J.;
RT   "Molecular Cloning and Characterization of the doxA Cytochrome P-450
RT   Hydroxylase Gene in Streptomyces peucetius subsp. caesius ATCC
RT   27952.";
RL   J. Microbiol. Biotechnol. 0:0-0(2001).
DR   EMBL; AF403708; AAK95626.1; -.
SQ   SEQUENCE   420 AA;  45645 MW;  2083C9FE9BA802C6 CRC64;


alignment_scores:
            Quality: 2062.00              Length:    422
              Ratio:  4.993                Gaps:      2
 Percent Similarity:  97.867   Percent Identity: 94.313

alignment_block:
 US-09-673-254-1/rev x Q93MI2   ..

 Align seg 1/1  to: Q93MI2  from: 1  to: 420

    1351 GTGAGCGGCGAGGCGCCCCGGGTGGCCGTCGACCCGTTCGCGTGTCCCAT 1302
         :::|||||||||||||||||||||||||||||||||||||||||||||||
       1 MetSerGlyGluAlaProArgValAlaValAspProPheAlaCysProMe 17

    1301 GATGACCATGCAGCGCAAGCCCGAGGTGCACGACGCCTTCCGGGAGGCGG 1252
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tMetThrMetGlnArgLysProGluValHisAspAlaPheArgGluAlaG 34

    1251 GCCCGGTCGTCGAGGTGAACGCCCCCGCGGGCGGACCCGCCTGGGTCATC 1202
         ||||||||||||||||||||||||||||||||||||||||||   |||
      34 lyProValValGluValAsnAlaProAlaGlyGlyProAlaTrpPheIle 50

    1201 ACCGATGACGCCCTCGCCCGCGAGGTGCTGGCCGATCCCCGGTTCGTGAA 1152
         ||||||||||||||:::|||   |||||||||||||||:::|||||
      51 ThrAspAspAlaLeuSerArgTyrValLeuAlaAspProArgLeuValLy 67

    1151 GGACCCCGACCTCGCCCCCGCCGCCTGGCGGGGGGTGGACGACGGTCTCG 1102
         ||||||||||||||||||||||||||||||||   ||||||||||
      67 sAspProAspLeuAlaProAlaAlaTrpArgGlyValValAspGlyLeuA 84

    1101 ACATCCCCGTTCCGGAGCTGCGTCCGTTCACGCTCATCGCCGTGGACGGC 1052
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 spIleProValProGluLeuArgProPheThrLeuIleAlaValAspGly 100

    1051 GAGGCCCACCGGCGCCTGCGCCGCATCCACGCACCTGCGTTCAACCCGCG 1002
         ||||||||||||||||||:::|||||||||||||||||||||||||||||
       101 GluAlaHisArgArgLeuHisArgIleHisAlaProAlaPheAsnProAr 117

    1001 CCGGCTGGCCGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGC 952
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
117 gArgLeuAlaGluArgThrAspArgIleAlaAlaIleAlaGlyArgLeuL 134

951 TCACCGAACTCGCCGACGCCTCCGGCCGGTCGGGCAAACCGGCCGAGCTG 902
    ||||||||||||||||||||||||||||||||||||||:::|||||||||||
134 euThrGluLeuAlaAspAlaSerGlyArgSerGlyGluProAlaGluLeu 150

901 ATCGGCGGCTTCGCGTACCACTTCCCGCTGTTGGTCATCTGCGAGCTGCT 852
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 IleGlyGlyPheAlaTyrHisPheProLeuLeuValIleCysGluLeuLe 167

851 CGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGCCGTCAGCGTTC 802
    ||||||||||||||||   |||   ||||||||||||||||||||||||
167 uGlyValProValThrValPro...MetAlaArgGluAlaValSerValL 183

801 TCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCACG 752
    ||||||||||:::   ::::::   ||||||||||||||||||||||||
183 euLysAlaLeuAla...SerAlaAlaGlnSerGlyGlyGlyAspGlyThr 198

751 GACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGCCTGCTCCT 702
    ||||||||||||||||||||||||||||||||||||||||||||||||||
199 AspProAlaGlyGlyValProAspThrSerAlaLeuGluSerLeuLeuLe 215

701 CGAAGCCGTGCACTCAGCCCGGCGGAACGACACCCCGACCATGACCCGCG 652
    ||||||||||||||||||||||||||||||||||||||||||||||||||
215 uGluAlaValHisSerAlaArgArgAsnAspThrProThrMetThrArgV 232

651 TGCTGTACGAGCGCGCGCAGGCCGAGTTCGGCTCGGTCTCCGACGACCAG 602
    ||||||||||:::::|||||||||||||||||||||||||||:::|||
232 alLeuTyrGluHisThrGlnAlaGluPheGlySerValSerAspAsnGln 248

601 CTCGTCTACATGATCACCGGGCTCATCTTCGCCGGCCACGACACCACCGG 552
    ||||||||||||||||||:::|||||||||||||||||:::   |||||
249 LeuValTyrMetIleThrGlyIleIlePheAlaGlyHisGluArgThrGl 265

551 CTCCTTCCTGGGCTTCCTGCTCGCGGAGGTCCTGGCGGGCCGCCTCGCGG 502
    ||||||||||||||||||||||||||||||||||||||||||||||||||
265 ySerPheLeuGlyPheLeuLeuAlaGluValLeuAlaGlyArgLeuAlaA 282

501 CGGATGCCGACGAGGACGCCGTCTCCCGGTTCGTGGAGGAGGCGCTGCGC 452
    ||||||||||||||||||||||||||||||||||||||||||||:::|||
282 laAspAlaAspGluAspAlaValSerArgPheValGluGluAlaValArg 298

451 TACCACCCGCCGGTGCCCTACACGTTGTGGAGGTTCGCTGCCACGGAGGT 402
    ||||||||||||||||||||||||||||||||||||||||||||||||||
299 TyrHisProProValProTyrThrLeuTrpArgPheAlaAlaThrGluVa 315

401 GACCATCGGCGGCGTCCGGCTGCCCCGCGGAGCGCCGGTGCTGGTGGACA 352
    ||||||||||||||||||||||   ||||||||||||||||||||||||
315 lThrIleGlyGlyValArgLeuProProGlyAlaProValLeuValAspI 332

351 TCGAGGGCACCAACACCGACGGCCGCCATCACGACGCCCCGCACGCCTTC 302
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 32 leGluGlyThrAsnThrAspGlyArgHisHisAspAlaProHisAlaPhe 348

 01 CACCCGGACCGTCCCTCGTGGCGGCGGCTCACCTTCGGCGACGGGCCGCA 252
    ||||||||||||||||||||||||||||||||||||||||||||||||||
349 HisProAspArgProSerTrpArgArgLeuThrPheGlyAspGlyProHi 365

251 CTACTGCATCGGGGAGCAGCTCGCCCAGCTGGAGTCGCGCACGATGATCG 202
    ||||||||||||||||||||||||||||||||||||||||||||||||||
365 sTyrCysIleGlyGluGlnLeuAlaGlnLeuGluSerArgThrMetIleG 382

201 GCGTACTGCGCAGCAGGTTCCCCGAGGCCCGACTGGCCGTGCCGTACGAC 152
    ||||||||||||||||||||||||||||||||||||||||||||||||||
382 lyValLeuArgSerArgPheProGluAlaArgLeuAlaValProTyrAsp 398

151 GAGTTGCGGTGGTGCCGGAAGGGGGCCCAGACGGCGCGGCTCACCGAACT 102
    ||||||||||||||||||:::|||||||||||||||||||||||||||||
399 GluLeuArgTrpCysArgAsnGlyAlaGlnThrAlaArgLeuThrGluLe 415

101 GCCCGTCTGGCTGCGC 86
    ||||||||||||||||
415 uProValTrpLeuArg 420
```

```
seq_name: sp_bacteria:Q9ZAU2

seq_documentation_block:
ID   Q9ZAU2      PRELIMINARY;      PRT;   275 AA.
AC   Q9ZAU2;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DOXORUBICIN BIOSYNTHESIS ENZYME DNRV.
GN   DNRV.
OS   Streptomyces peucetius.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1950;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29050;
RX   MEDLINE=99084971; PubMed=9864344;
RA   Lomovskaya N., Otten S.L., Doi-Katayama Y., Fonstein L., Liu X.-C.,
RA   Takatsu T., Inventi A., Filippi S., Torti F., Colombo A.L.,
RA   Hutchinson C.R.;
RT   "Doxorubicin overproduction in Streptomyces peucetius: cloning and
RT   characterization of the dnrU ketoreductase and dnrV genes and the doxA
RT   cytochrome P-450 hydroxylase gene.";
RL   J. Bacteriol. 181:305-318(1999).
DR   EMBL; U77891; AAD04716.1; -.
DR   InterPro; IPR004360; Gly_bleo_diox.
DR   Pfam; PF00903; Glyoxalase; 2.
SQ   SEQUENCE   275 AA;  28481 MW;  DEE13C0E714D4397 CRC64;


alignment_scores:
            Quality: 1465.00              Length:    275
              Ratio:  5.327                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.636

alignment_block:
 US-09-673-254-1/rev x Q9ZAU2   ..

 Align seg 1/1  to: Q9ZAU2  from: 1  to: 275

    2175 GTGACCAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCC 2126
         :::||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetThrArgPheAlaProGlyAlaProAlaTrpPheAspLeuGlySerPr 17

    2125 CGATGTCGCCGCCTCGGCCGACTTCTACACCGGCCTGTTCGGCTGGACCG 2076
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 oAspValAlaAlaSerAlaAspPheTyrThrGlyLeuPheGlyTrpThrA 34

    2075 CCACCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACGTTCAGCTCC 2026
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 laThrValValSerAspProGlyAlaGlyGlyTyrThrThrPheSerSer 50

    2025 GACGGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCCCTA 1976
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AspGlyLysLeuValAlaAlaValAlaArgHisGlnIleAspThrProTy 67

    1975 CCACCGGCCGTACGGGCCCGGGAACGACCAGCACGGCATGCCGGCCATCT 1926
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rHisArgProTyrGlyProGlyAsnAspGlnHisGlyMetProAlaIleT 84

    1925 GGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGGGTC 1876
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 rpThrValTyrPheAlaThrAspAspAlaAspAlaLeuThrLysArgVal 100

    1875 GAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGGCCT 1826
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GluThrAlaGlyGlyGluValIleMetThrProMetAspValLeuGlyLe 117

    1825 CGGCCGGATGGCGGTCTTCGCCGACCCCGCCGGGGCCGCGTTCGCGGTCT 1776
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uGlyArgMetAlaValPheAlaAspProAlaGlyAlaAlaPheAlaValT 134
```

```
    1775 GGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGCTCG 1726
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rpArgLysGlyValMetGluGlyAlaGluValThrGlyValProGlySer 150

    1725 GTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCCGGGACTT 1676
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 ValGlyTrpValGluLeuValThrAspGlyIleGlyAlaAlaArgAspPh 167

    1675 CTACCCGGCGACCCTCGGCCTGGCTCCGGCCGACACCGGACTGAAGGGCG 1626
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eTyrProAlaThrLeuGlyLeuAlaProAlaAspThrGlyLeuLysGlyV 184

    1625 TCACCGACCCGGTCTGGCACATCGGTGACACACCGGTCGCCGGCACCCAG 1576
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 alThrAspProValTrpHisIleGlyAspThrProValAlaGlyThrGln 200

    1575 GAGCTGGGCGTCACCGGCGCGGTACGGCCGCACTGGGCCGTGCTGTTCGC 1526
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GluLeuGlyValThrGlyAlaValArgProHisTrpAlaValLeuPheAl 217

    1525 CGTGCACGACTGCGACGCGACGGTCCGGCGCGCCGTTGAACTCGGCGGCT 1476
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 aValHisAspCysAspAlaThrValArgArgAlaValGluLeuGlyGlyS 234

    1475 CCGTCGAGAACGAGCCCGCCGACACGCCCAGGGGGCGGCGGGCGGACCTG 1426
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 erValGluAsnGluProAlaAspThrProArgGlyArgArgAlaAspLeu 250

    1425 CTCGACCCGCACGGGGCCGGCTTCTCGGTGGTCGAACTGCGGGAGGGGTA 1376
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuAspProHisGlyAlaGlyPheSerValValGluLeuArgGluGlyTy 267

    1375 CCCCGCGGCGGCGGGCGGTGCCTCG 1351
         |||||||||||||||||||||||||
     267 rProAlaAlaAlaGlyGlyAlaSer 275

seq_name: sp_bacteria:Q55078

seq_documentation_block:
ID   Q55078      PRELIMINARY;      PRT;   275 AA.
AC   Q55078;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
     ORFA PROTEIN.
     ORFA.
     Streptomyces sp.
     Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
     Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1931;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C5;
RX   MEDLINE=96236067; PubMed=8655530;
RA   Dickens M.L., Strohl W.R.;
RT   "Isolation and characterization of a gene from Streptomyces sp. strain
RT   C5 that confers the ability to convert daunomycin to doxorubicin on
RT   Streptomyces lividans TK24.";
RL   J. Bacteriol. 178:3389-3395(1996).
DR   EMBL; U50973; AAB08048.1; -.
DR   InterPro; IPR004360; Gly_bleo_diox.
DR   Pfam; PF00903; Glyoxalase; 2.
SQ   SEQUENCE   275 AA;  28840 MW;  BCDD9BEA60BDE4AF CRC64;


alignment_scores:
            Quality: 1385.00              Length:    275
              Ratio:  5.149                Gaps:      0
 Percent Similarity:  97.818   Percent Identity: 94.182

alignment_block:
 US-09-673-254-1/rev x Q55078   ..
```

```
Align seg 1/1  to: Q55078  from: 1  to: 275

    2175 GTGACCAGGTTCGCGCCCGGCGCCCCCGCATGGTTCGACCTCGGTTCGCC 2126
         :::||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetThrArgPheAlaProGlyAlaProAlaTrpPheAspLeuGlySerPr 17

    2125 CGATGTCGCCGCCTCGGCCGACTTCTACACCGGCCTGTTCGGCTGGACCG 2076
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 oAspValAlaAlaSerAlaAspPheTyrThrGlyLeuPheGlyTrpThrA 34

    2075 CCACCGTGGTCAGCGACCCGGGCGCCGGGGGATACACGACGTTCAGCTCC 2026
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 laThrValValSerAspProGlyAlaGlyGlyTyrThrThrPheSerSer 50

    2025 GACGGGAAGCTGGTCGCCGCGGTCGCCCGCCACCAGATCGACACCCCCTA 1976
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AspGlyLysLeuValAlaAlaValAlaArgHisGlnIleAspThrProTy 67

    1975 CCACCGGCCGTACGGGCCCGGGAACGACCAGCACGGCATGCCGGCCATCT 1926
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rHisArgProTyrGlyProGlyAsnAspGlnHisGlyMetProAlaIleT 84

    1925 GGACCGTGTACTTCGCCACCGACGACGCCGACGCACTGACCAAGCGGGTC 1876
         ||||||||||||||||||:::||||||||||||||||||||||||||||
      84 rpThrValTyrPheAlaThrAsnAspAlaAspAlaLeuThrLysArgVal 100

    1875 GAGACGGCGGGCGGCGAGGTCATCATGACTCCGATGGACGTCCTCGGCCT 1826
         ||||||||||||||:::||||||||||||||||||||||||||||||||
     101 GluThrAlaGlyGlyAspValIleMetThrProMetAspValLeuGlyLe 117

    1825 CGGCCGGATGGCGGTCTTCGCCGACCCCGCCGGGGCCGCGTTCGCGGTCT 1776
         ||||||||||||||||||||||||||:::|||||||||||||||||||||
     117 uGlyArgMetAlaValPheAlaAspProSerGlyAlaAlaPheAlaValT 134

    1775 GGCGCAAGGGAGTCATGGAGGGCGCGGAGGTGACGGGCGTGCCCGGCTCG 1726
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rpArgLysGlyValMetGluGlyAlaGluValThrGlyValProGlySer 150

    1725 GTCGGCTGGGTCGAGCTGGTGACCGACGGCATCGGGGCCGCCCCGGGACTT 1676
         ||||||||||||||||||||||||||   ||||||:::||||||   ||
     151 ValGlyTrpValGluLeuValThrAspAspIleGlyThrAlaArgGlyPh 167

    1675 CTACCCGGCGACCCTCGGCCTGGCTCCGGCCGACACCGGACTGAAGGGCG 1626
         ||||   |||||||||||||||||||||||||||||||   |||||||
     167 eTyrArgAlaThrLeuGlyLeuAlaProAlaAspThrGlyArgLysGlyV 184

    1625 TCACCGACCCGGTCTGGCACATCGGTGACACACCGGTCGCCGGCACCCAG 1576
         ||||||||||||||||||||   ||||||||||||||||||||:::
     184 alThrAspProValTrpHisIleHisAspThrProValAlaGlyThrArg 200

    1575 GAGCTGGGCGTCACCGGCGCGGTACGGCCGCACTGGGCCGTGCTGTTCGC 1526
         ||||||||:::||||||||||||||||||||||||||||||||||||::
     201 GluLeuGlyThrThrGlyAlaValArgProHisTrpAlaValLeuPheSe 217

    1525 CGTGCACGACTGCGACGCGACGGTCCGGCGCGCCGTTGAACTCGGCGGCT 1476
         :||||||||||||||||||||||||||||||||||||||||||||||||||
     217 rValHisAspCysAspAlaThrValArgArgAlaValGluLeuGlyGlyS 234

    1475 CCGTCGAGAACGAGCCCGCCGACACGCCCAGGGGGCGGCGGGCGGACCTG 1426
         ||||||||||||||||:::||||||||||||||||||||||||||||||||
     234 erValGluAsnGluProValAspThrProArgGlyArgArgAlaAspLeu 250

    1425 CTCGACCCGCACGGGGCCGGCTTCTCGGTGGTCGAACTGCGGGAGGGGTA 1376
         ||||||||||||||||||||||||||||||||||||||||||||||:::||
     251 LeuAspProHisGlyAlaGlyPheSerValValGluLeuArgGluAlaTy 267

    1375 CCCCGCGGCGGCGGGCGGTGCCTCG 1351
         ||||||||||||   |||||||||
     267 rProAlaAlaAlaAspGlyAlaSer 275

seq_name: sp_bacteria:Q9ZAU1
```

```
seq_documentation_block:
ID   Q9ZAU1      PRELIMINARY;      PRT;   287 AA.
AC   Q9ZAU1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DAUNORUBICIN C-13 KETOREDUCTASE.
GN   DNRU.
OS   Streptomyces peucetius.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1950;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29050;
RX   MEDLINE=99084971; PubMed=9864344;
RA   Lomovskaya N., Otten S.L., Doi-Katayama Y., Fonstein L., Liu X.-C.,
RA   Takatsu T., Inventi A., Filippi S., Torti F., Colombo A.L.,
RA   Hutchinson C.R.;
RT   "Doxorubicin overproduction in Streptomyces peucetius: cloning and
RT   characterization of the dnrU ketoreductase and dnrV genes and the doxA
RT   cytochrome P-450 hydroxylase gene.";
RL   J. Bacteriol. 181:305-318(1999).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; U77891; AAD04717.1; -.
DR   InterPro; IPR002198; ADH_short.
DR   Pfam; PF00106; adh_short; 1.
KW   Oxidoreductase.
SQ   SEQUENCE   287 AA;  30632 MW;  4F9448EAE55F6191 CRC64;

alignment_scores:
            Quality: 1014.00              Length:    196
              Ratio:  5.173                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-673-254-1/rev x Q9ZAU1   ..

 Align seg 1/1  to: Q9ZAU1  from: 1  to: 287

    2869 CGGTACCCGCGCATCGATGTCATGGCCAGCAACGCCGGCGGCATGTTCTG 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      92 ArgTyrProArgIleAspValMetAlaSerAsnAlaGlyGlyMetPheTr 108

    2819 GTCGCGCACCACGACCCAGGACGGGTTCGAGGCCACCATCCAGGTCAATC 2770
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     108 pSerArgThrThrThrGlnAspGlyPheGluAlaThrIleGlnValAsnH 125

    2769 ACCTCGCAGGCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGC 2720
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      25 isLeuAlaGlyPheLeuLeuAlaArgLeuLeuArgGluArgLeuAlaGly 141

      19 GGGCGGCTGATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGGATCGA 2670
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     142 GlyArgLeuIleLeuThrSerSerAspAlaTyrThrGlnGlyArgIleAs 158

    2669 CCCGGACGACCTCAACGGCGACCGTCACCGCTACAGCGCCGGCCAGGCGT 2620
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     158 pProAspAspLeuAsnGlyAspArgHisArgTyrSerAlaGlyGlnAlaT 175

    2619 ACGGCACGTCCAAACAGGCCAACATCATGACCGCGGCGGAGGCCGCCAGG 2570
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     175 yrGlyThrSerLysGlnAlaAsnIleMetThrAlaAlaGluAlaAlaArg 191

    2569 CGCTGGCCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTCCGCAC 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     192 ArgTrpProAspValLeuAlaValSerTyrHisProGlyGluValArgTh 208

    2519 CCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCAACCCCT 2470
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     208 rArgIleGlyArgGlyThrValAlaSerSerTyrPheArgPheAsnProP 225

    2469 TCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTGGCGTCC 2420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      225 heLeuArgSerAlaAlaLysGlyAlaAspThrLeuValTrpLeuAlaSer 241

    2419 GCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGACCGGCGGCT 2370
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      242 AlaProAlaGluGluLeuThrThrGlyGlyTyrTyrSerAspArgArgLe 258

    2369 GTCCCCGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCT 2320
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      258 uSerProValSerGlyProThrAlaAspAlaGlyLeuAlaAlaLysLeuT 275

    2319 GGGAGGCCGGCGCGGCCGCCGTCGGCGACACCGCGCAC 2282
         |||||||||||||||||||||||||||||||||||||
      275 rpGluAlaGlyAlaAlaAlaValGlyAspThrAlaHis 287

seq_name: sp_bacteria:Q53877

seq_documentation_block:
ID   Q53877      PRELIMINARY;      PRT;   287 AA.
AC   Q53877;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KETOREDUCTASE.
OS   Streptomyces sp. C5.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=45212;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C5;
RX   MEDLINE=96236066; PubMed=8655529;
RA   Dickens M.L., Ye J., Strohl W.R.;
RT   "Cloning, sequencing, and analysis of aklaviketone reductase from
RT   Streptomyces sp. strain C5.";
RL   J. Bacteriol. 178:3384-3388(1996).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; U43704; AAB08016.1; -.
DR   InterPro; IPR002198; ADH_short.
DR   Pfam; PF00106; adh_short; 1.
KW   Oxidoreductase.
SQ   SEQUENCE   287 AA;  30799 MW;  8FEB9D11BAFE1E78 CRC64;

alignment_scores:
            Quality:  980.00              Length:    195
              Ratio:  5.026                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 96.410

alignment_block:
 US-09-673-254-1/rev x Q53877   ..

 Align seg 1/1  to: Q53877  from: 1  to: 287

    2869 CGGTACCCGCGCATCGATGTCATGGCCAGCAACGCCGGCGGCATGTTCTG 2820
         ||||||||||||||||||||||||||:::||||||||||||||||||||
      92 ArgTyrProArgIleAspValMetAlaGlyAsnAlaGlyGlyMetPheTr 108

    2819 GTCGCGCACCACGACCCAGGACGGGTTCGAGGCCACCATCCAGGTCAATC 2770
         ||||||||||||||||||||||||||||||||||:::||||||||||||
     108 pSerArgThrThrThrGlnAspGlyPheGluAlaThrLeuGlnValAsnH 125

    2769 ACCTCGCAGGCTTCCTGCTGGCACGGCTGCTGCGGGAGCGGCTCGCGGGC 2720
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     125 isLeuAlaGlyPheLeuLeuAlaArgLeuLeuArgGluArgLeuAlaGly 141

    2719 GGGCGGCTGATCCTCACCTCGTCCGACGCGTACACCCAGGGCCGGATCGA 2670
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     142 GlyArgLeuIleLeuThrSerSerAspAlaTyrThrGlnGlyArgIleAs 158
```

```
2669 CCCGGACGACCTCAACGGCGACCGTCACCGCTACAGCGCCGGCCAGGCGT 2620
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 158 pProAspAspLeuAsnGlyAspArgHisArgTyrSerAlaGlyGlnAlaT 175

2619 ACGGCACGTCCAAACAGGCCAACATCATGACCGCGGCGGAGGCCGCCAGG 2570
     ||||||||||||||||||||||||||||||||||:::||||||||||||
 175 yrGlyThrSerLysGlnAlaAsnIleMetThrAlaThrGluAlaAlaArg 191

2569 CGCTGGCCGGACGTGCTGGCGGTCAGCTATCACCCCGGTGAGGTCCGCAC 2520
     ||||||||||||||||||:::||||||||||||||||||||||||||||
 192 ArgTrpProAspValLeuThrValSerTyrHisProGlyGluValArgTh 208

2519 CCGCATCGGACGGGGCACGGTCGCCTCGTCCTACTTCCGGTTCAACCCCT 2470
     ||||||||||||||||||||||||||:::||||||||||||||||||||
 208 rArgIleGlyArgGlyThrValAlaSerThrTyrPheArgPheAsnProP 225

2469 TCCTGCGCTCCGCGGCGAAGGGCGCCGACACCCTCGTGTGGCTGGCGTCC 2420
     ||||||||||||||||||||||||||||||||||||||||||||||:::
 225 heLeuArgSerAlaAlaLysGlyAlaAspThrLeuValTrpLeuAlaAla 241

2419 GCGCCGGCCGAGGAGTTGACCACGGGCGGCTACTACAGCGACCGGCGGCT 2370
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 242 AlaProAlaGluGluLeuThrThrGlyGlyTyrTyrSerAspArgArgLe 258

2369 GTCCCCGGTGAGCGGCCCGACCGCCGACGCCGGCCTCGCGGCGAAGCTCT 2320
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 258 uSerProValSerGlyProThrAlaAspAlaGlyLeuAlaAlaLysLeuT 275

2319 GGGAGGCCGGCGCGGCCGCCGTCGGCGACACCGCG 2285
     ||||||||:::|||||||||||||||||||||||
 275 rpGluAlaSerAlaAlaAlaValGlyAspThrAla 286
```